

Máster en Estadística e Investigación Operativa

Título: SEGMENTACIÓN DE CLIENTES: FIDELIDAD, "SHARE OF WALLET" Y SENSIBILIDAD AL PRECIO

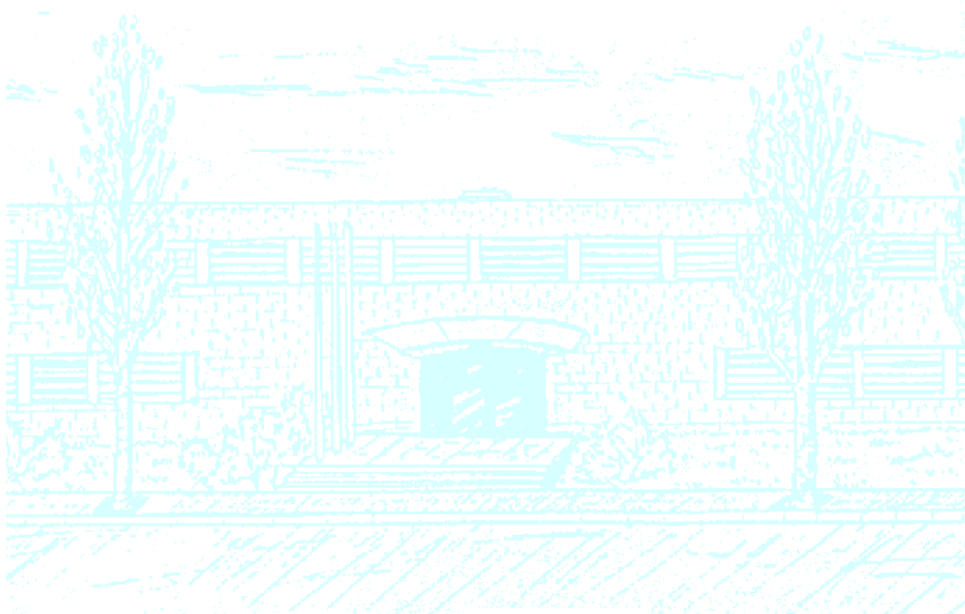
Autor: Miguel Rico Paino

Directora: Monica Becue

Tutor de empresa: Pau Agulló (Neo Metrics)

Departamento: Estadística e Investigación Operativa

Convocatoria: Junio 2011



**Facultat de Matemàtiques
i Estadística**

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya

TRABAJO FIN DE MASTER

Master en Estadística e Investigación Operativa

Segmentación de Clientes: fidelidad, "Share of Wallet" y sensibilidad al precio

Miguel Rico Paino

Directora: Mònica Bécue Bertaut
Tutor de empresa: Pau Agulló (Neo Metrics)

Departamento de estadística e Investigación Operativa
Universidad Politécnica de Cataluña

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a la Obra social de la Fundación “la Caixa” y más concretamente a su programa de becas para másteres en España, la oportunidad que me han brindado de poder estudiar un Master en Estadística e Investigación operativa.

Ha sido un privilegio durante estos dos años haber sido becario de la Fundación “la Caixa” ya que realmente he disfrutado aprendiendo nuevas técnicas estadísticas y matemáticas, lo que ha supuesto un gran complemento especial a mi formación.

También he de agradecer especialmente el apoyo de Dra.Mònica Bécue Bertaut sobre todo por sus valiosos consejos a la hora de desarrollar este proyecto, a Pau Agulló por haberme dado la oportunidad de haber realizado este proyecto y a todos los compañeros de la oficina de Neo-Metrics por todo lo que me han enseñado.

Además me gustaría reconocer la labor de mis padres, hermana, familiares y amigos por todo su apoyo y comprensión mostrada siempre.



RESUMEN

En este proyecto, realizado bajo un convenio de colaboración para la empresa consultora Neo-Metrics, se estudian y analizan las características y los hábitos de compras de los consumidores de una cadena de supermercados ubicado en Cataluña.

El proyecto consiste básicamente en realizar diferentes segmentaciones de clientes. Primero se segmenta a los clientes dependiendo su grado de fidelidad. Después se realiza una segmentación en función del “Share of Wallet” y finalmente se obtienen grupos homogéneos de clientes en función de su sensibilidad al precio. Además se aplicará a los datos de compra de los clientes el algoritmo del Market Basket Analysis para analizar reglas de asociación entre los diferentes productos que son comprados por los consumidores.

Palabras clave: Análisis de Correspondencias Múltiples, Análisis factoriales, Análisis cluster, Market Basket Analysis, K-Means, Cluster jerárquico.

MSC2000: 62H30, 62H25, 68T10

ABSTRACT

This project, which has been conducted under a collaboration agreement for the consulting firm Neo-Metrics, analyzes consumer buying habits in a supermarket chain located in Catalonia.

The project performs various customer segmentations. First, customers are divided into different groups according to their degree of loyalty. Second, customers are clustered based upon their “Share of Wallet”. Finally, they are divided into homogeneous groups of customers according to their price of sensitivity.

Note that the Market Basket Analysis algorithm is applied in order to analyze association rules between the different products that are purchased by consumers.

Keywords: Multiple correspondence analysis, Factor analysis, Cluster analysis, K-means, Hierarchical clustering, Market Basket Analysis, Pattern recognition.

MSC2000: 62H30, 62H25, 68T10

Índice General

1.INTRODUCCIÓN	11
1.1ORIGEN DEL PROYECTO: LOS CUPONES DE DESCUENTO	12
1.2 OBJETIVOS Y DESCRIPCIÓN DEL PROYECTO	13
1.3 UTILIDAD DEL PROYECTO	14
2.INFORMACIÓN DISPONIBLE	15
2.1 TARJETA DE FIDELIZACIÓN	15
2.2 BENEFICIOS DE UNA TARJETA DE FIDELIZACIÓN	15
2.3 ESTRUCTURA DE LA BASE DE DATOS	16
2.4 INFORMACIÓN SOBRE SECCIONES	19
2.5 CREACIÓN DE VARIABLES	20
2.5.1 PORCENTAJE DE REDENCIÓN	20
2.5.2 GASTO MENSUAL	21
2.5.3 NÚMERO DE COMPRAS GRANDES	21
2.5.4 TIPOLOGÍA DE PRODUCTOS	23
2.5.5 ÍNDICE DE SENSIBILIDAD AL PRECIO	24
2.6 TRAMIFICACIÓN DE VARIABLES	25
2.6.1 TRAMIFICACIÓN DE VARIABLES DE FIDELIDAD	25
2.6.2 TRAMIFICACIÓN DE GASTO MENSUAL EN CADA SECCIÓN	25
2.6.3 TRAMIFICACIÓN DEL ÍNDICE DE SENSIBILIDAD	27
3.ANÁLISIS BIDIMENSIONAL.....	28
4.SEGMENTACIÓN EXPERTA	31
5.ESTUDIO DE LA REDENCIÓN	32
5.1 CARACTERIZACIÓN GENERAL DE LOS CLIENTES	32
5.2 GRUPO SIN SENSIBILIDAD AL CUPÓN	34
5.3 GRUPO CON LA MÁXIMA SENSIBILIDAD AL CUPÓN (Redención>10%)	35
5.4 GRUPO CON UNA SENSIBILIDAD AL CUPÓN REDUCIDA	36
5.5 GRUPO CON UNA SENSIBILIDAD AL CUPÓN MEDIA.....	37
5.6 GRUPO CON UNA SENSIBILIDAD AL CUPÓN ALTA	38
6.TÉCNICAS DE DATA MINING EMPLEADAS	39

6.1 ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES.....	40
6.2 CLASIFICACIÓN AUTOMÁTICA	43
6.2.1 ALGORITMOS DE CLASIFICACIÓN	44
6.2.1.1 MÉTODOS DE PARTICIÓN DIRECTA(K-MEANS)	45
6.2.1.2 CLASIFICACIÓN ASCENDENTE JERÁRQUICA	46
7.SEGMENTACIÓN POR FIDELIDAD	47
7.1 DESCRIPCIÓN DE LA SEGMENTACIÓN.....	47
7.2METODOLOGÍA	48
7.3RESULTADOS.....	49
7.3.1 RESULTADOS ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES	49
7.3.2 RESULTADOS SEGMENTACIÓN CON ALGORITMO K-MEANS.....	58
7.3.2.1 DESCRIPCIÓN SEGMENTO 1: “Los más fieles”	61
7.3.2.2 DESCRIPCIÓN SEGMENTO 2 “Los que más visitan”	62
7.3.2.3 DESCRIPCIÓN SEGMENTO 3 :“Gasto alto, clientes más selectivos”	63
7.3.2.4 DESCRIPCIÓN SEGMENTO 4 :“Gasto medio con redención media”	64
7.3.2.5 DESCRIPCIÓN SEGMENTO 5 “Los más sensibles al cupón”	65
7.3.2.6 DESCRIPCIÓN SEGMENTO 6 “Gasto medio con escasa sensibilidad al cupón”	66
7.3.2.7 DESCRIPCIÓN SEGMENTO 7 “Los menos fieles”	67
7.3.3 RESULTADOS SEGMENTACIÓN CON ALGORITMO JERÁRQUICO	68
7.3.3.1 DESCRIPCIÓN SEGMENTO 1 “Los más fieles”	72
7.3.3.2 DESCRIPCIÓN SEGMENTO 2 “Grado de fidelidad alto pero no compran en todas las secciones”	73
7.3.3.3 DESCRIPCIÓN SEGMENTO 3 “Gasto medio sin sensibilidad a los cupones	74
7.3.3.4 DESCRIPCIÓN SEGMENTO 4 “Gasto medio con mucha sensibilidad a los cupones”	75
7.3.3.5 DESCRIPCIÓN SEGMENTO 5 “Gasto medio con un nivel de redención medio alto”	76
7.3.3.6 DESCRIPCIÓN SEGMENTO 6 “Gasto medio-bajo con escasa sensibilidad”	77
7.3.3.7 DESCRIPCIÓN SEGMENTO 7 “Los menos fieles”	78
8.SEGMENTACIÓN DEL “SHARE OF WALLET”	79
8.1DESCRIPCIÓN DE LA SEGMENTACIÓN.....	79
8.2METODOLOGÍA	81
8.3RESULTADOS.....	82
8.3.1RESULTADOS ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES	82
8.3.2 RESULTADOS SEGMENTACIÓN “SHARE OF WALLET”	90
8.3.2.1 DISTRIBUCIÓN GENERAL SEGMENTACIÓN “SHARE OF WALLET”	94
8.3.2.2.DESCRIPCIÓN SEGMENTO 1 :”No compran en la sección carnicería, charcutería, frutería, pescadería ni lácteos.....	97
8.3.2.3 DESCRIPCIÓN SEGMENTO 2 :”Solo compran ocasionalmente en la sección alimentación seca, pescadería y frutería”.	98
8.3.2.4 DESCRIPCIÓN SEGMENTO 3: “Compran todo en el supermercado excepto en la sección carnicería”	99
8.3.2.5 ESCRIPCIÓN SEGMENTO 4: “Compras ocasionales en todas las secciones excepto en carnicería”	100
8.3.2.6 DESCRIPCIÓN SEGMENTO 5 :”Solo compran ocasionalmente en la sección alimentación seca”	101
8.3.2.7DESCRIPCIÓN SEGMENTO 6: “Realizan compras ocasionales en todas las secciones”.	102
8.3.2.8 DESCRIPCIÓN SEGMENTO 7:: “Compras ocasionales excepto en pescadería, carnicería ,charcutería , frutería y perfumería”	103

8.3.2.9 DESCRIPCIÓN SEGMENTO 8: "No compran en la sección carnicería ni charcutería pero en promedio gastan más y tienen un nivel de redención más alto"	104
8.3.2.10 DESCRIPCIÓN SEGMENTO 9 : "Solo compran ocasionalmente en la sección alimentación seca, conservas y líquidos"	105
8.3.2.11 DESCRIPCIÓN SEGMENTO 10 : "Compras grandes excepto en carnicería, charcutería, frutería, pescadería y perfumería"	106
8.3.2.12 DESCRIPCIÓN SEGMENTO 11: "No compran en la sección carnicería ni charcutería"	107
8.3.2.13 DESCRIPCIÓN SEGMENTO 12: "Los que compran todo en el supermercado"	108
9.SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO	109
9.1 DESCRIPCIÓN DE LA SEGMENTACIÓN	109
9.2 METODOLOGÍA DE SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO A NIVEL GENERAL	111
9.3 RESULTADOS SENSIBILIDAD AL PRECIO A NIVEL GENERAL	112
9.3.1 DISTRIBUCIÓN GENERAL	115
9.3.2 SEGMENTO 1: "Preferencia de productos de gama media"	117
9.3.3 SEGMENTO 2: "Preferencia por productos baratos"	118
9.3.4 SEGMENTO 3: "Compran por igual productos baratos, caros y productos de precio medio"	119
9.3.5 SEGMENTO 4: "Preferencia por productos caros"	120
9.4 METODOLOGÍA DE SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO A NIVEL SECCIÓN	121
9.5 RESULTADOS	123
9.5.1 ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES	123
9.5.2.1 DISTRIBUCIÓN GENERAL	136
9.5.2.2 SEGMENTO 1 "Los menos sensibles al precio: tendencia a adquirir productos caros"	139
9.5.2.3 SEGMENTO 2 "Sensibilidad media en todas las secciones excepto en lácteos"	140
9.5.2.4 SEGMENTO 3 "Sensibilidad media en todas las secciones"	141
9.5.2.5 SEGMENTO 4 "Sensibilidad media excepto en congelado y lácteos"	142
9.5.2.6 SEGMENTO 5 "Los más sensibles al precio"	143
9.5.2.7 SEGMENTO 6 "No compra en la sección congelados y sensibilidad baja en perfumería"	144
9.5.2.8 SEGMENTO 7 "Sensibilidad muy alta en perfumería"	145
9.5.2.9 SEGMENTO 8 "Sensibilidad muy alta en congelado y media en las demás secciones"	146
9.5.2.10 SEGMENTO 9 "El segundo grupo que menos gasto realiza"	147
9.5.2.11 SEGMENTO 10: "El segmento menos fiel"	148
10. MARKET BASKET ANALYSIS	149
10.1 ¿QUÉ ES EL MARKET BASKET ANALYSIS?	149
10.2 OBJETIVOS	152
10.3 APLICACIÓN EN LOS DATOS	152
10.3.1 PRIMEROS RESULTADOS	153
10.3.2 RESULTADOS POSTERIORES	155
10.4 CARACTERIZACIÓN DE REGLAS DE DECISIÓN	157
10.5 CONCLUSIONES DEL MBA	159
11.CONCLUSIONES GENERALES	160

BIBLIOGRAFIA	162
---------------------------	------------

1.INTRODUCCIÓN

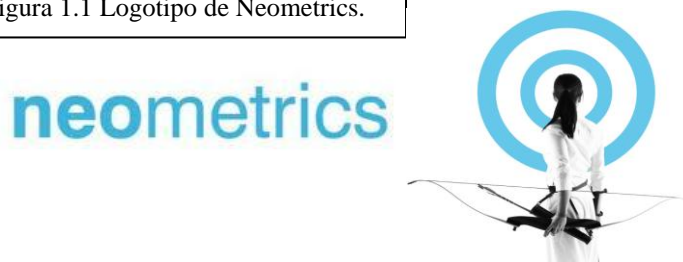
Este proyecto final del Master de Estadística e Investigación Operativa (MEIO) está realizado como colaboración con la empresa Neometrics en un proyecto para un supermercado ubicado en Cataluña que se enmarca dentro del marco de las relaciones Universidad-Empresa.

Le empresa Neometrics es una consultora internacional especializada en el análisis de datos o “Inteligencia analítica” para la optimización de los procesos de toma decisiones empresariales.

La compañía, con capital 100% español, fue fundada en 2003 por un grupo de profesionales de diferentes disciplinas (economía, matemáticas, estadística e informática) que decidieron utilizar las matemáticas y la estadística para aplicarlo a problemas reales que tienen las empresas.

Además en 2007 Neo Metrics recibió el premio KDD Cup, la más prestigiosa competición de análisis de datos en el ámbito internacional, siendo la primera empresa europea en conseguirlo.

Figura 1.1 Logotipo de Neometrics.



Las soluciones de Neo Metrics abarcan desde sistemas de previsión de la demanda, análisis de redes sociales, modelos de detección del fraude y sobre todo la gestión y optimización de campañas de Marketing y sistemas de recomendación.

Por motivos de confidencialidad la empresa cliente para la cual se hace el proyecto la denominamos “**Supermercado**”.

1.1 ORIGEN DEL PROYECTO: LOS CUPONES DE DESCUENTO

Este proyecto fin de master surge como un proyecto complementario a otro proyecto que está llevando a cabo Neo Metrics para la empresa “Supermercado” y que consiste en crear un sistema automático y personalizado de recomendaciones de cupones de descuento a clientes con el objetivo de aumentar el número de visitas y el gasto de los clientes. Antes de explicar los objetivos de este proyecto es imprescindible saber qué es un cupón de descuento.

Figura 1.1.1

¿Qué es un cupón de descuento?

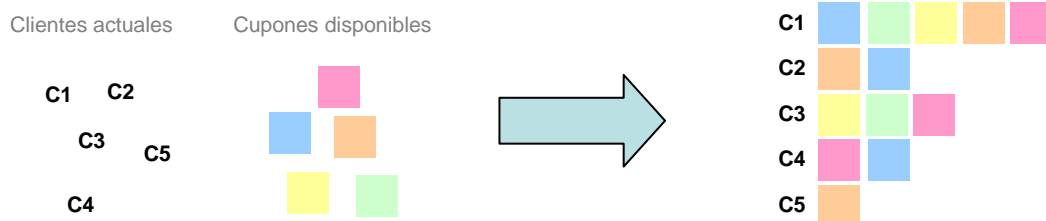
Un cupón de descuento es un material impreso que el supermercado entrega al consumidor por el cual este puede obtener un descuento en un producto determinado si canjea o redime el cupón en una compra posterior. Los cupones de descuento se suelen utilizar como una herramienta de marketing para aumentar la fidelización del consumidor y fomentar las visitas.



El problema principal de este proyecto, que se esquematiza en la figura 1.1.2, se basa en la idea de generar una asignación óptima de cupones a clientes dada una lista de cupones y una lista de clientes maximizando el número de cupones redimidos, las visitas y el gasto por cliente.

El sistema no está enfocado a maximizar la redención únicamente, sino a maximizar el impacto económico del cliente a partir de la redención del cupón. El cupón es un medio, no un fin en sí mismo.

Figura 1.1.2 Esquema explicativo del objetivo del proyecto.



Maximizando número de redenciones, visitas y/o gasto

Para esta asignación de cupones resulta relevante segmentar a los clientes e identificar clientes sensibles al precio, clientes con una alta fidelidad al supermercado y clientes con un alto nivel de sensibilidad a los cupones. Además será importante estudiar reglas de asociación entre diferentes productos detectando correlaciones entre los diferentes productos que son adquiridos en el supermercado.

1.2 OBJETIVOS Y DESCRIPCIÓN DEL PROYECTO

El proyecto se centra en el tema del “Customer Intelligence”, es decir, es un proyecto que trata de obtener e inferir patrones de comportamiento de los clientes analizando en detalle el comportamiento de éstos con el **objetivo de poder alterar su comportamiento de compra mediante estrategias de marketing.**

Los objetivos del proyecto se pueden resumir en estos cuatro apartados:

1) Segmentación de los clientes en función de su grado de fidelidad.

Se pretende identificar segmentos de clientes que tienen un alto grado de fidelidad. Entendemos que un consumidor tiene un grado fidelidad alto cuando tiene un gasto mensual elevado, visita con cierta frecuencia el supermercado y realiza la mayor parte de sus compras en el mismo establecimiento.

Dentro de esta segmentación es importante distinguir segmentos de clientes que son **sensibles al uso de los cupones de descuento que ofrece el supermercado.**

Además se efectuará un retrato robot caracterizando a cada segmento con variables sociodemográficas que ayuden a interpretar y comprender mejor cada clase.

2) Segmentación de los clientes en función de su “Share of Wallet”.

Se quiere segmentar a los clientes en función de cómo reparten su gasto dentro de las diferentes secciones del “Supermercado”. Más concretamente se pretende detectar grupos de clientes que tienden a comprar la totalidad de su compra en el mismo establecimiento y segmentos de clientes que no realizan toda su compra en el mismo establecimiento, sino que escogen diferentes supermercados para realizar su compra. Para este tipo de clientes se pretende investigar en qué secciones realizan compras considerables y en qué secciones su gasto es nulo.

También se caracterizará a los diferentes grupos con variables sociodemográficas que ayuden a interpretar y comprender mejor cada segmento.

3) Segmentación de los clientes en función de la sensibilidad al precio.

El objetivo planteado es segmentar a los clientes en función de su sensibilidad al precio o tendencia de adquirir productos baratos. Se quiere detectar diferentes tipos de clientes:

- Segmentos de clientes que tienen preferencia por productos baratos.
- Segmentos de clientes que tienen preferencia por productos de precio medio.
- Segmentos de clientes que tienen preferencia por productos caros.

Además se pretende detectar grupos de clientes que tienden a comprar productos caros en unas secciones determinadas pero que suelen comprar productos baratos en otras. Por ejemplo, puede existir un segmento de consumidores que compran productos baratos en todas las secciones excepto en la sección perfumería en la que optan por comprar productos caros.

4) Segmentación por producto: Análisis de la cesta de la compra.

El objetivo es aplicar el algoritmo del “Basket Market Analysis” para estudiar asociaciones o correlaciones entre los diferentes productos que son comprados por los consumidores.

Se trata de realizar un análisis de la cesta de la compra que descubrirá un conjunto de reglas de asociación entre productos.

Software utilizado y lenguaje de programación:

El software utilizado para la realización de este proyecto ha sido fundamentalmente SAS Enterprise Guide y SAS Miner. No se tenía conocimientos de programación en SAS y gracias a la realización de este proyecto se ha conseguido tener un nivel medio-alto de programación en SAS.

También se ha utilizado el paquete multivariante SPAD para aplicar técnicas multivariantes.

1.3 UTILIDAD DEL PROYECTO

Como se ha comentado en el apartado 1.1 de origen del proyecto, este proyecto que presento forma parte de otro proyecto real para un sistema automático y personalizado de recomendaciones de cupones de descuento a clientes. Por lo tanto, este proyecto de forma indirecta ayudará a Neo-Metrics y al “supermercado” (cliente) a la consecución del objetivo de modificar el comportamiento de compra de ciertos segmentos de clientes aumentando su gasto y su número de visitas. Para ello es necesario segmentar a los clientes e identificar clientes sensibles al precio, clientes con una alta fidelidad al supermercado y clientes con un alto nivel de sensibilidad a los cupones de descuento.

Además este proyecto proporciona información muy interesante para la alta dirección del supermercado o el departamento de Marketing. La identificación de segmentos de clientes fieles, sensibles al precio y sensibles al cupón puede ser una herramienta clave para realizar un plan de fidelización de Marketing y modificar el comportamiento de compra de ciertos segmentos mediante determinadas acciones de Marketing. Lo que se propone es que toda decisión empresarial esté respaldada por un análisis estadístico exhaustivo.

También el análisis de asociación entre productos puede apoyar a la hora de tomar ciertas decisiones estratégicas como la ubicación de productos, la creación de packs promocionales y una política de descuentos más eficaz.

2.INFORMACIÓN DISPONIBLE

2.1 TARJETA DE FIDELIZACIÓN

Se parte de una base de datos de clientes procedente de una tarjeta de fidelización de clientes. Básicamente se dispone de información del histórico de compras y del uso de cupones del cliente desde el año 2009 hasta enero de 2011. Se disponen de 180 millones de registros de compras realizadas por más de 200,000 clientes. Además solo se tiene información de aquellas compras que han sido efectuadas por clientes que han utilizado su tarjeta de fidelización a la hora de efectuar el pago de sus compras.

Figura 2.1.1 Tarjeta de fidelización.



2.2 BENEFICIOS DE UNA TARJETA DE FIDELIZACIÓN

El beneficio para el consumidor es claro: beneficiarse de descuentos o promociones cuando efectúe alguna compra mediante el uso de cupones.

Una ventaja para el supermercado que emite la tarjeta de fidelidad es la posibilidad de conocer hábitos de consumo de sus clientes y, en consecuencia, orientar la política comercial y de marketing de la compañía.

Para obtener la tarjeta, es necesario rellenar un impreso con los datos del usuario. Al realizar compras y cargarlas en la tarjeta, se pueden cruzar los datos del comprador con los productos adquiridos. De este modo, por medio de técnicas estadísticas o de Data Mining es posible analizar el comportamiento de los diferentes clientes.

Las empresas emisoras de tarjetas, además de fidelizar a sus clientes habituales consiguen obtener patrones de comportamiento sobre los diferentes clientes que resultan útiles para sus campañas de Marketing.

Además se puede disponer de información de compras que pueden resultar útil para la toma de decisiones estratégicas.

2.3 ESTRUCTURA DE LA BASE DE DATOS

La base de datos de la que se parte es una base de datos relacional que se estructura en 13 tablas. Al trabajar con 13 tablas es necesario realizar consultas y crear nuevas tablas. Para manipular la información se ha aprendido los fundamentos del lenguaje SQL: lenguaje de consulta estructurado. Como se comentaba en el apartado anterior se ha tenido que manejar tablas de un gran volumen de datos (180 millones de compras para más de 200,000 clientes).

La información disponible tal como se muestra en la figura 2.3.1 se puede dividir en cuatro grandes apartados:



Para cada apartado se presentan las variables más relevantes que se han utilizado a lo largo de este proyecto. Muchas de estas variables han sido creadas a partir de otras mediante la unión de varias tablas.

-Información sociodemográfica del cliente.

Figura 2.3.2 Información sociodemográfica del cliente.	
Perfil sociodemográfico del cliente	Variables sociodemográficas
	Fecha nacimiento
	Sexo
	Idioma (Catalán o Castellano)
	Número de familiares en el hogar
	Número de menores de edad en el hogar
	Antigüedad del cliente
	Edad del cliente

Como se puede ver en la tabla superior tenemos información referente al perfil sociodemográfico del cliente. Estas variables son importantes para caracterizar a los diferentes segmentos que se obtendrán en las diferentes segmentaciones.

Algunas de estas variables han sido categorizadas en diferentes modalidades como el número de familiares en el hogar, el número de menores de edad en el hogar, la antigüedad del cliente y la edad.

-Información sobre el uso de cupones.

Figura 2.3.3 Información sobre el uso de cupones.	
Uso de cupones	Información uso de cupones
	Número de cupones recibidos
	Número de cupones redimidos
	Porcentaje de redención
	Fecha de redención

El uso de cupones es una dimensión clave en este proyecto ya que a partir de ella se puede inferir la sensibilidad al uso de los cupones de cada cliente.

-Información de la dimensión ventas.

Figura 2.3.4 Información de variables relacionadas con la dimensión ventas.	
Ventas	Información ventas
	Precio Venta al Público (PVP)
	Importe total de cada producto
	Importe oferta por producto
	Forma de pago: efectivo o tarjeta de crédito
	Gasto mensual.
	Número de visitas mensuales.
	Número de compras grandes.
	Número de productos totales comprados durante el 2010 por cada cliente.
	Número de productos baratos comprados durante el 2010.
	Número de productos de gama media comprados durante el 2010.
	Número de productos caros comprados durante el 2010.
	Porcentaje de productos caros comprados respecto al total.
	Porcentaje de productos de precio medio comprados respecto al total.
	Porcentaje de productos baratos comprados respecto al total.
	Gasto mensual en cada sección*.
	Porcentaje de productos de marca blanca comprados.
	Porcentaje de productos adquiridos que no son marca blanca.
	Índice de sensibilidad.
	Índice de sensibilidad para cada sección.

De la dimensión ventas se obtienen variables que serán utilizadas posteriormente en la segmentación de fidelidad: gasto mensual, número de visitas y número de compras grandes. Además destacan variables relacionadas con la sensibilidad al precio como son porcentaje de productos caros adquiridos, porcentaje de productos de gama media, porcentaje de productos baratos, índice de sensibilidad e índice de sensibilidad para cada sección. Esta última variable hace referencia al índice de sensibilidad de cada cliente en las 12 secciones consideradas, por lo que se trata de 12 variables.

Para la segmentación del “Share of Wallet” se ha creado la variable gasto mensual en cada sección. Esta variable hace referencia al gasto de cada cliente en cada sección del supermercado y por lo tanto se desglosa en 12 variables.

-Información de las acciones o promociones y de los artículos.

Figura 2.3.5 Información de variables relacionadas con acciones, descuentos y tipología de productos.

Acciones, descuentos y tipología de productos	Información Acciones y Artículos
	Descripción de artículos.
	Marca.
	Formato(kg,liros,unidades,...).
	Tipología de producto:barato,precio medio o caro.
	Marca Blanca: si/no.
	Tipo de descuento.
	Sección a la que pertenece.
	Descuento aplicado.
	Regalos.
	Tipo de cupón de descuento.

Con estas tablas se puede obtener información relativa a la tipología de productos (marca blanca, producto barato, caro o de precio medio), a la sección a la que pertenece cada artículo y sobre todo al tipo de descuento u oferta comercial (regalos, descuentos porcentuales por producto y otras promociones).

Tanto la variable tipología de producto como marca blanca han sido creadas a partir de otras variables.

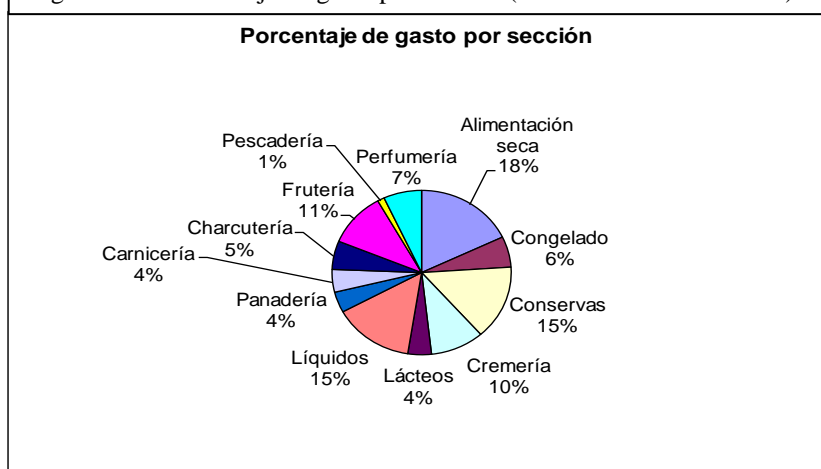
La variable descripción del producto será utilizada en el apartado del Market Basket Analysis como variable target para descubrir asociaciones entre productos.

2.4 INFORMACIÓN SOBRE SECCIONES

Como forma de organización, los supermercados estructuran sus productos en diferentes secciones. Dentro de cada sección existen otras subsecciones donde hay diferentes categorías de producto.

El siguiente gráfico (gráfico 2.4.1) muestra la distribución del gasto por secciones. Como se desprende del gráfico la sección más importante es alimentación seca ya que representa aproximadamente el 18% del gasto total para el período comprendido entre enero de 2009 y octubre de 2010. Además cabe destacar el fuerte peso que tiene la sección frutería y cremería.

Figura 2.4.1 Porcentaje de gasto por sección (enero 2009-octubre 2010).



Para entender un poco mejor de qué productos están compuestas las diferentes subsecciones, se presentan las siguientes tablas en la que se detallan los productos que forman parte de cada una de ellas.

Tabla 2.4.2 Categorías de productos en Alimentación seca.

CATEGORÍAS DE PRODUCTO ALIMENTACIÓN SECA	
ACEITES	CAFÉS Y SUCEDÁNEOS
ALIMENTOS DIETÉTICOS	CARAMELOS Y GOLOSINAS
ALIMENTOS INFANTILES	CEREALES Y BARRITAS CEREALES
ALIMENTOS NAVIDEÑOS	CHOCOLATES Y CACAOS
ALIMENTOS Y PRODUCTOS PARA ANIMALES	ESPECIAS Y CONDIMENTOS
APERITIVOS	FRUTOS SECOS ESTUCHADOS
AZUCAR	GALLETAS
HARINAS, PASTAS Y SEMOLAS	SOPAS, CREMAS, POTAJES, PURES
HUEVOS	TÉS E INFUSIONES
LEGUMBRES Y ARROZ	TOMATES FRITOS Y SALSAS
MERMELADAS, MIEL, MEMBRILLO	VINAGRES Y ADEREZANTES
POSTRES A PREPARAR	PRODUCTOS INTERNACIONALES
PRODUCTOS ECOLÓGICOS	SAL Y BICARBONATO

Tabla 2.4.3 Categorías de productos en la sección conservas y líquidos

CATEGORÍAS DE PRODUCTO: CONSERVAS	CATEGORÍAS DE PRODUCTO: LÍQUIDOS
CONSERVAS DE ACEITUNAS	AGUAS Y GASEOSAS
CONSERVAS DE CARNE	CAVAS, ESPUMOSOS I SIDRAS
CONSERVAS DE ENCURTIDOS	CERVEZAS
CONSERVAS DE FRUTA	LICORES
CONSERVAS DE PESCADO	REFRESCOS E ISOTONICOS
CONSERVAS DE PLATOS PREPARADOS	VINOS
CONSERVAS VEGETALES	ZUMOS Y BEBIDAS DE FRUTAS

Tabla 2.4.4 Categorías de productos en la sección cremería.

CATEGORÍAS DE PRODUCTO: CREMERÍA	
LECHE FRESCA	QUESO DE UNTAR
MANTEQUILLAS	QUESO FRESCO PLS.
MARGARINAS	QUESO FUNDIDO, RALLADO, AZUL.
NATAS	SALSAS FRESCAS
PASTAS FRESCAS	SOPAS Y CREMAS REFRIGERADAS
POSTRES	YOGHOURTS
	ZUMOS Y BEBIDAS FRESCAS

2.5 CREACIÓN DE VARIABLES

En este apartado se explicará brevemente cómo se han creado ciertas variables que resultan clave para la consecución de los objetivos planteados.

2.5.1 PORCENTAJE DE REDENCIÓN

La variable propensión de redención hace referencia al uso del cupón y se expresa matemáticamente de la siguiente forma:

$$\text{Porcentaje de redención} = \frac{\text{Número de cupones redimidos por cliente}}{\text{Número de cupones recibidos por cliente}}$$

2.5.2 GASTO MENSUAL

La variable gasto mensual y número de visitas mensuales se calcula como el cociente entre el gasto total de un cliente en el período considerado y el número de meses considerados.

$$\text{Gasto mensual} = \frac{\text{Gasto total período considerado}}{\text{Número de meses}}$$

2.5.3 NÚMERO DE COMPRAS GRANDES

A continuación se describe la formación de esta variable. En primer lugar se explica el supuesto que se ha considerado para la creación de esta variable.

Se considera que un cliente compra realmente en una sección o realiza una “compra grande” en una sección cuando su gasto mensual en promedio es superior o igual a los umbrales mostrados en la siguiente tabla (Tabla 2.5.3.1):

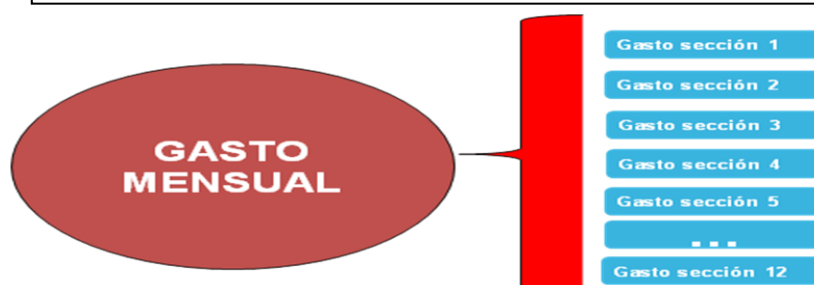
Por ejemplo cuando un cliente tiene un promedio en el gasto mensual superior a 10€ consideramos que realmente compra en la sección alimentación seca. Esta tabla ha sido validada por el departamento de Marketing de nuestro cliente.

Tabla 2.5.3.1	
SECCIÓN	Gasto mensual
Alimentación seca	10 €
Congelado	5 €
Conservas	8 €
Cremeria	5 €
Lácteos	3 €
Líquidos	8 €
Pan	3 €
Carnicería	5 €
Charcutería	5 €
Frutas	8 €
Pescadería	3 €
Perfumería	4 €

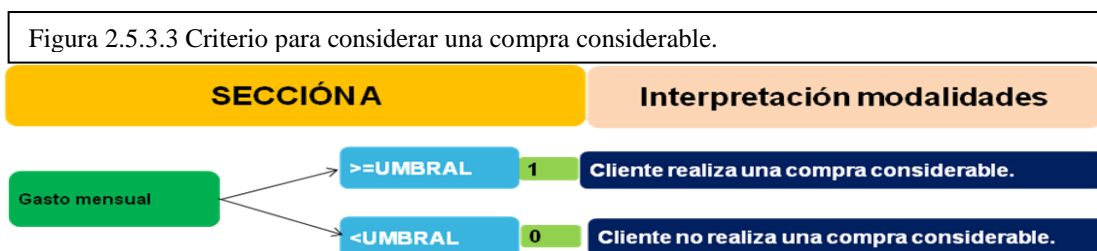
A partir del supuesto anterior se construye la variable siguiendo estas 3 etapas:

- 1) Se calcula el gasto mensual que corresponde a cada una de las 12 secciones.

Figura 2.5.3.2 Desglose del gasto mensual



- 2) A partir de la tabla con los umbrales (tabla 2.5.3.1) clasificamos a los clientes en dos tipologías:
- Clientes que realizan una compra considerable cuando su gasto mensual en la sección determinada es mayor o igual al umbral establecido en la tabla 2.5.3.1
 - Clientes que no realizan una compra considerable cuando su gasto mensual en la sección determinada es inferior al umbral establecido en la tabla 2.5.3.1



- 3) Se suman todas las compras grandes que ha realizado cada cliente en todas las secciones. Según esto, el número de compras grandes es una variable que toma valores enteros cuyo valor máximo es 12, ya que consideramos 12 secciones. Si un cliente realiza 12 compras grandes esto significa que este cliente realiza un gasto considerable en todas las secciones y su grado de fidelidad es muy alto. El valor mínimo de esta variable es 0. En este caso se dice que el cliente no ha realizado ninguna compra considerable en ninguna sección.

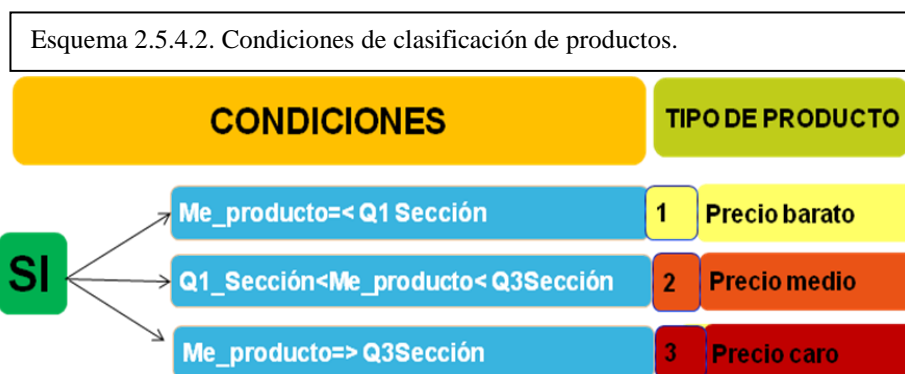
2.5.4 TIPOLOGÍA DE PRODUCTOS

En este subapartado mostramos el procedimiento utilizado para clasificar a los productos de una forma automática y así obtener la **variable tipología de producto**. El proceso es el siguiente:

- 1) En primer lugar para poder comparar entre productos tenemos que tener precios unitarios. Por lo tanto dividiremos el precio del producto entre el número de unidades calculando así el precio unitario de cada producto.
- 2) Después se calcula el precio mediano de cada producto, el precio mediano de la sección a la que pertenece el producto (Sección Leche Entera), el precio del primer cuartil y el del tercer cuartil de la sección a la que pertenece.

IDARTICULO	Sección 10	Me Producto	Me Sección 10	Q1 Sección 10	Q3 Sección 10	Tipo Producto
1690	LECHE ENTERA BRIK	0.89 €	0.79 €	0.77 €	0.85 €	3
1710	LECHE ENTERA BRIK	0.79 €	0.79 €	0.77 €	0.85 €	2
4252	LECHE ENTERA BRIK	0.68 €	0.79 €	0.77 €	0.85 €	1
4428	LECHE ENTERA BRIK	1.70 €	0.79 €	0.77 €	0.85 €	3
13374	LECHE ENTERA BRIK	0.85 €	0.79 €	0.77 €	0.85 €	3
13888	LECHE ENTERA BRIK	0.79 €	0.79 €	0.77 €	0.85 €	2
17886	LECHE ENTERA BRIK	1.01 €	0.79 €	0.77 €	0.85 €	3
32727	LECHE ENTERA BRIK	0.69 €	0.79 €	0.77 €	0.85 €	1
46215	LECHE ENTERA BRIK	0.85 €	0.79 €	0.77 €	0.85 €	3
47771	LECHE ENTERA BRIK	0.85 €	0.79 €	0.77 €	0.85 €	3
49835	LECHE ENTERA BRIK	0.95 €	0.79 €	0.77 €	0.85 €	3
53513	LECHE ENTERA BRIK	0.51 €	0.79 €	0.77 €	0.85 €	1

- 3) A partir de la tabla anterior establecemos condiciones para clasificar los productos en tres categorías.



Según el esquema cuando el precio mediano del producto es inferior o igual al precio del primer cuartil de la sección a la que pertenece, entonces clasificamos a ese producto en la categoría de productos baratos.

Cuando el precio mediano del producto se encuentra comprendido entre el precio del primer cuartil y el precio del tercer cuartil de la sección a la que pertenece, clasificamos a ese producto como un producto de precio medio.

Cuando el precio mediano del producto sea superior o igual al precio del tercer cuartil de la sección a la que pertenece, el producto es clasificado como un producto de precio caro.

En el caso de que el precio mediano del producto coincida con el precio mediano de la sección a la que pertenece, con el precio del primer cuartil y el del tercer cuartil de la sección de la sección a la que pertenece, el producto es clasificado como un producto de precio medio.

2.5.5 ÍNDICE DE SENSIBILIDAD AL PRECIO

Como se explicará en los siguientes capítulos la creación de un índice de sensibilidad de cada cliente ayudará a identificar segmentos de clientes que son sensibles en unas secciones pero insensibles en otras. Por lo tanto se calculará un índice de sensibilidad al precio para todas las secciones.

Para la construcción de este índice se ha de calcular la proporción de productos baratos, caros y de gama media respecto al total de productos adquiridos para cada cliente.

Esquema 2.5.5.1. Variables para construir el índice de sensibilidad.							
COMPRAS AÑO 2010							
Id Cliente	Número de productos comprados	Núm. Productos precio medio	Núm. Productos Baratos	Núm. Productos Caros	Proporción productos precio medio	Proporción productos caros	Proporción productos baratos
1001							
1002							
1003							

El índice de sensibilidad al precio se expresa de la siguiente forma en una escala porcentual:

$$I.Sensibilidad = \left[\left((1 \times \% \text{ productos } _ \text{ caros}) + (1,5 \times \% \text{ productos } _ \text{ gama } _ \text{ media}) + (2 \times \% \text{ productos } _ \text{ baratos}) \right) - 1 \right] \times 100$$

-Si el **Índice de sensibilidad=0** se dice que existe **sensibilidad nula al precio**, es decir, el cliente solo compra productos caros.

-Si el **Índice de sensibilidad=1** se supone que hay una sensibilidad máxima al precio, es decir, el cliente solo compra productos baratos.

2.6 TRAMIFICACIÓN DE VARIABLES

En este apartado se mostrará brevemente la transformación de variables continuas en variables ordinales.

2.6.1 TRAMIFICACIÓN DE VARIABLES DE FIDELIDAD

Esquema 2.6.1.1. Tramificación de variables de fidelidad.

VARIABLES DE FIDELIZACIÓN							
Redención		Visitas mensuales		Gasto mensual		Número de compras grandes	
Tramos	Valores	Tramos	Valores	Tramos	Valores	Tramos	Valores
1	0%	1	<1 visita	1	<20€	1	0
2	[0,3%)	2	[1-3) visitas	2	[20€-50€)	2	1-3
3	[3%-6%)	3	[3-5) visitas	3	[50€-100)	3	4-6
4	[6%-10%)	4	[5-7) visitas	4	[100€-200€)	4	7-9
5	>=10%	5	>=7 visitas	5	>=200€	5	>9

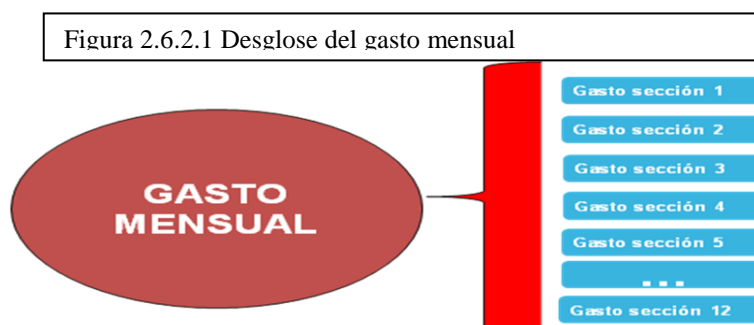
Como se observa en la tabla anterior se recodifican las variables de fidelidad en cinco tramos, siendo el primer tramo el que toma el valor mínimo y por lo tanto el que nos indica un nivel de fidelidad muy reducido. Estas variables serán utilizadas en la segmentación en función de fidelidad.

2.6.2 TRAMIFICACIÓN DE GASTO MENSUAL EN CADA SECCIÓN

El objetivo de esta recodificación es detectar clientes que realmente compran en una sección determinada y destinan una parte importante de su gasto en una sección. Como sucedía en el apartado 2.5.3 se considera que un cliente compra realmente en una sección o realiza una “compra grande” en una sección cuando su gasto mensual en promedio es superior o igual a los umbrales mostrados en la tabla 2.5.3.1.

El proceso de creación de esta variable y su tramificación es la siguiente:

- 1) En primer lugar y tal como muestra la siguiente figura se procede a desglosar el gasto mensual entre las doce secciones calculándose el gasto mensual que corresponde a cada sección.



- 2) Después de obtener el gasto mensual por cliente para cada sección se lleva a cabo la categorización del gasto mensual en cada sección tal como se indica en el siguiente esquema. Esta categorización se basa en el criterio de la tabla con los umbrales para cada sección (tabla 2.5.3.1).

Figura 2.6.2.2 Tramificación de la variable gasto mensual en cada sección



De acuerdo con el esquema anterior existen 3 modalidades diferentes:

-**Modalidad 1 o de gasto nulo:** aquellos clientes que nunca han comprado en esa sección.

-**Modalidad 2 o de gasto inferior al umbral:** nos indica clientes que compran en esa sección de forma puntual o que realizan compras ocasionales. El gasto mensual de estos clientes es inferior al umbral

-**Modalidad 3 o de gasto igual o superior al umbral:** los clientes compran realmente en una sección considerándose como un gasto importante en esa sección. El gasto mensual de estos clientes es igual o superior al umbral.

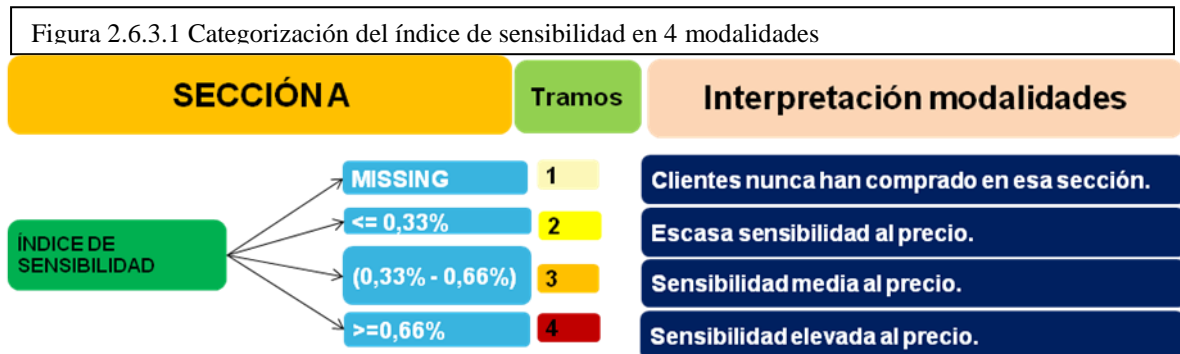
Para entender mejor esto se muestra como ejemplo la categorización del gasto mensual por cliente en la sección alimentación, en la sección congelado y en la sección perfumería.

Figura 2.6.2.3 Categorización del gasto mensual en alimentación, congelado y perfumería.

GASTO MENSUAL POR CLIENTE EN CADA SECCIÓN					
Gasto sección alimentación		Gasto sección congelado		Gasto sección perfumería	
Tramos	Valores	Tramos	Valores	Tramos	Valores
1	Gasto nulo	1	Gasto nulo	1	Gasto nulo
2	Gasto < 10€	2	Gasto < 5€	2	Gasto < 4€
3	Gasto >= 10€	3	Gasto >= 5€	3	Gasto >= 4€

2.6.3 TRAMIFICACIÓN DEL ÍNDICE DE SENSIBILIDAD

A continuación se muestra la tramificación del índice de sensibilidad mencionado anteriormente en 4 modalidades.



Tal como se indica en la figura anterior se distinguen 4 modalidades:

- 1) Los que tienen un gasto nulo en esa sección determinada y por lo tanto no han comprado en esa sección.
- 2) Los que tienen una preferencia por productos caros o una escasa sensibilidad al precio (sensibilidad inferior o igual al 33%).
- 3) Los que tienen una sensibilidad media con un índice comprendido entre 0,33% y 0,66%
- 4) Los que tienen preferencia por productos baratos o una alta sensibilidad al precio y que cuentan con un nivel de sensibilidad superior o igual al 66%.

3.ANÁLISIS BIDIMENSIONAL

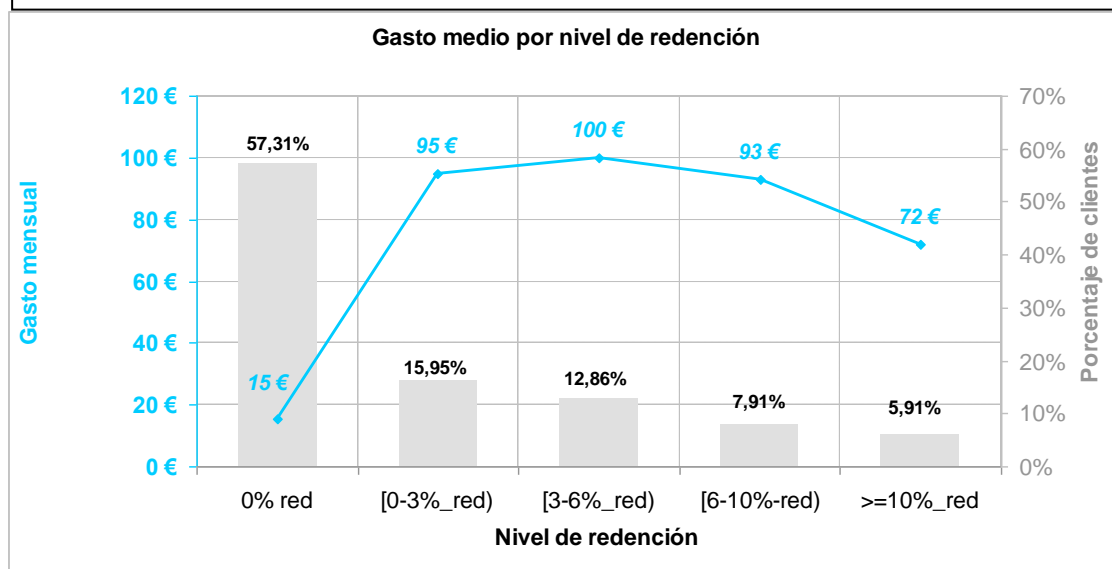
En este apartado se observará de forma general las relaciones más significativas entre las variables de fidelidad. Más concretamente se analizan las siguientes relaciones:

- Relación entre el nivel de redención y el gasto mensual.
- Relación entre el nivel de redención y el número de visitas.
- Relación entre el nivel de redención y el número de compras grandes
- Relación entre el nivel de visitas y el gasto mensual.

Este análisis se ha realizado con datos de enero de 2009 hasta octubre de 2010.

Relación entre el nivel de redención y el gasto mensual

Figura 3.1.Relación entre el gasto mensual en función del nivel de redención.

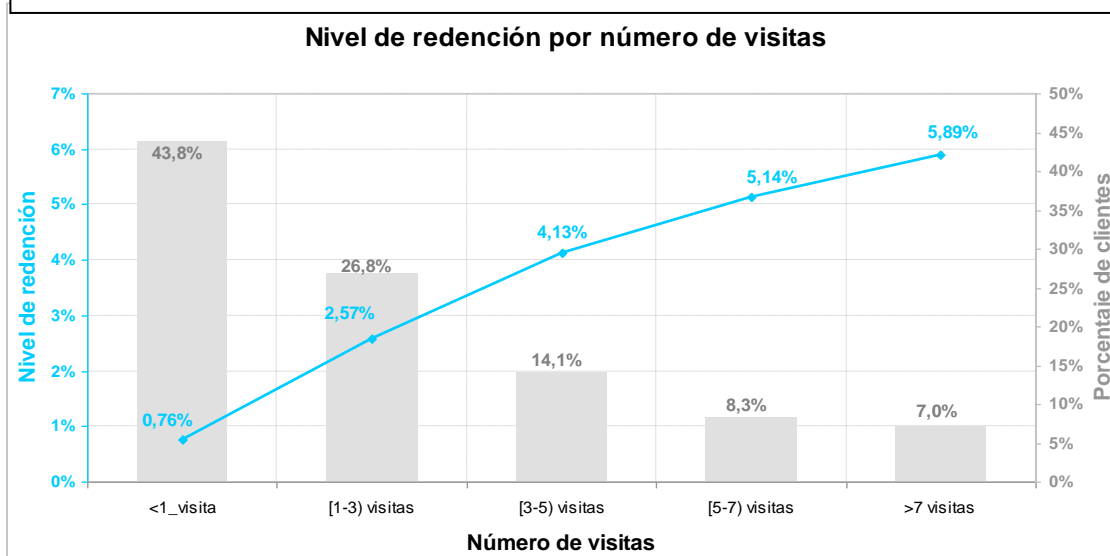


Existen muchas diferencias en términos de gasto entre el hecho de redimir y la acción de no redimir. El gasto medio de los clientes que no redimen es de 15.4€ en cambio aquellos clientes que tienen una tasa de redención baja [0- 3%) tienen un gasto medio de 95€.

Además se aprecia que a medida que aumenta la tasa de redención el gasto mensual medio va disminuyendo y que los clientes que tienen una alta tasa de redención gastan menos que aquellos que tienen una tasa media.

Relación entre el nivel de redención y el número de visitas

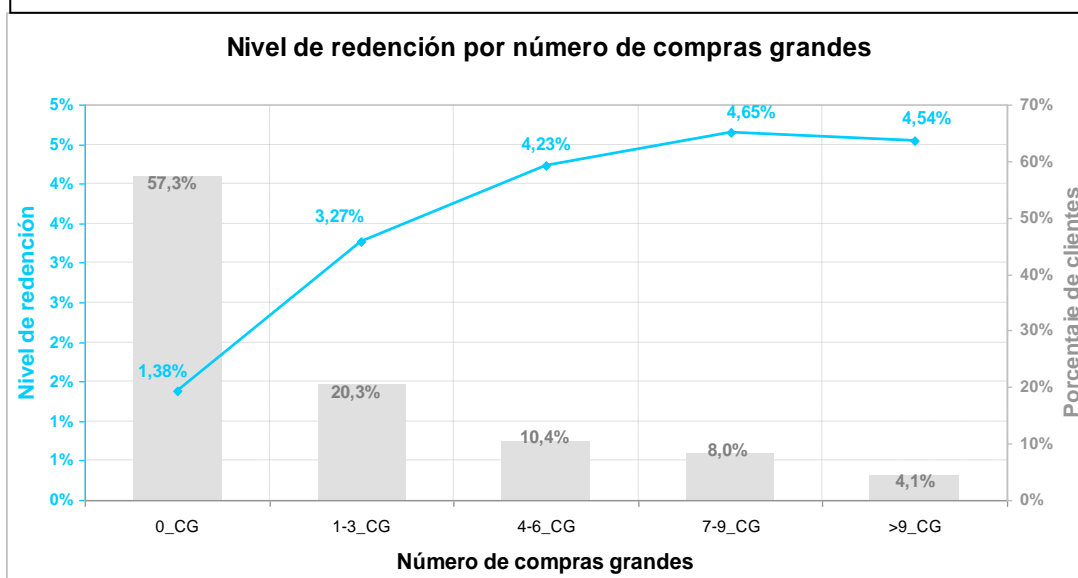
Figura 3.2. Relación entre el nivel de redención en función del número de visitas.



Además se aprecia que a medida que aumenta el número de visitas el nivel de redención aumenta. Existe una relación lineal claramente.

Relación entre el nivel de redención y el número de compras grandes

Figura 3.3. Relación entre el nivel de redención en función del número de compras grandes.

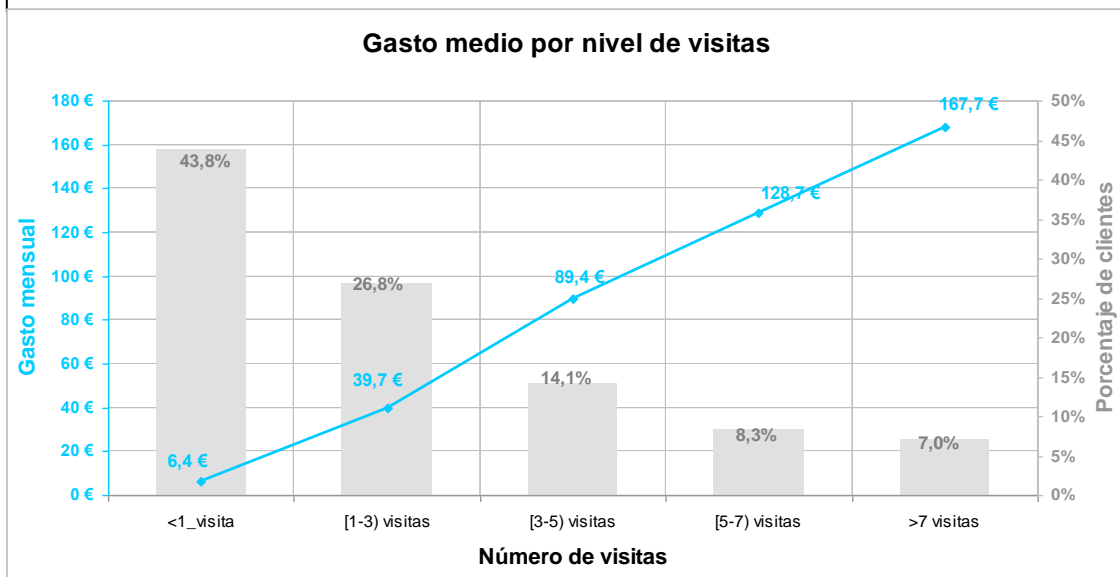


Aquellos que no realizan ninguna compra grande se caracterizan por tener un nivel de redención muy bajo. En cambio aquellos que realizan entre 1 y 3 compras grandes tienen un nivel de redención mucho más alto (superior al 3%).

El nivel de redención alcanza su máximo en el segmento de aquellos clientes que realizan entre 7 y 9 compras grandes. Para aquellos clientes que realizan casi la totalidad de su compra en el supermercado el nivel de redención disminuye ligeramente.

Relación entre el nivel de visitas y el gasto mensual

Figura 3.4. Relación entre el gasto medio en función del número de visitas.



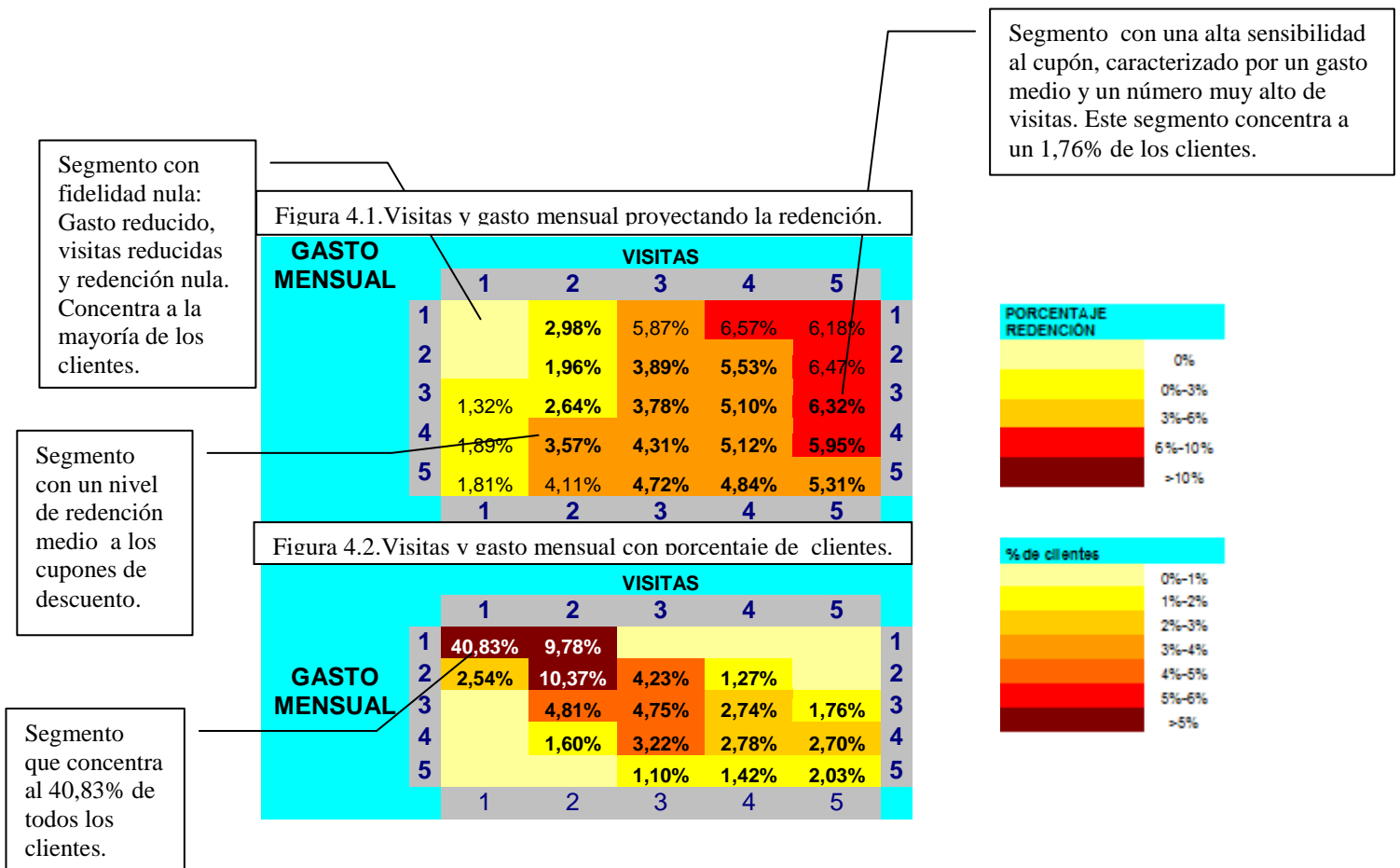
Lógicamente a medida que aumenta el número de visitas el gasto aumenta. Existe una relación lineal claramente.

4.SEGMENTACIÓN EXPERTA

Antes de efectuar la segmentación por fidelidad se ha realizado una segmentación experta en la que se segmenta a los clientes por su número de visitas al mes y su gasto mensual. Se categorizan las variables gasto mensual y número de visitas mensuales tal como se ha mostrado en los apartados anteriores. Como cada variable tiene 5 modalidades existen 25 posibles segmentos.

En el primer gráfico (figura 4.1) mostramos el cruce entre el número de visitas y el gasto mensual proyectando el porcentaje de redención que tiene cada segmento.

En el segundo gráfico (figura 4.2) mostramos el cruce entre el número de visitas y el gasto mensual pero mostrando el porcentaje de clientes que hay en cada segmento.



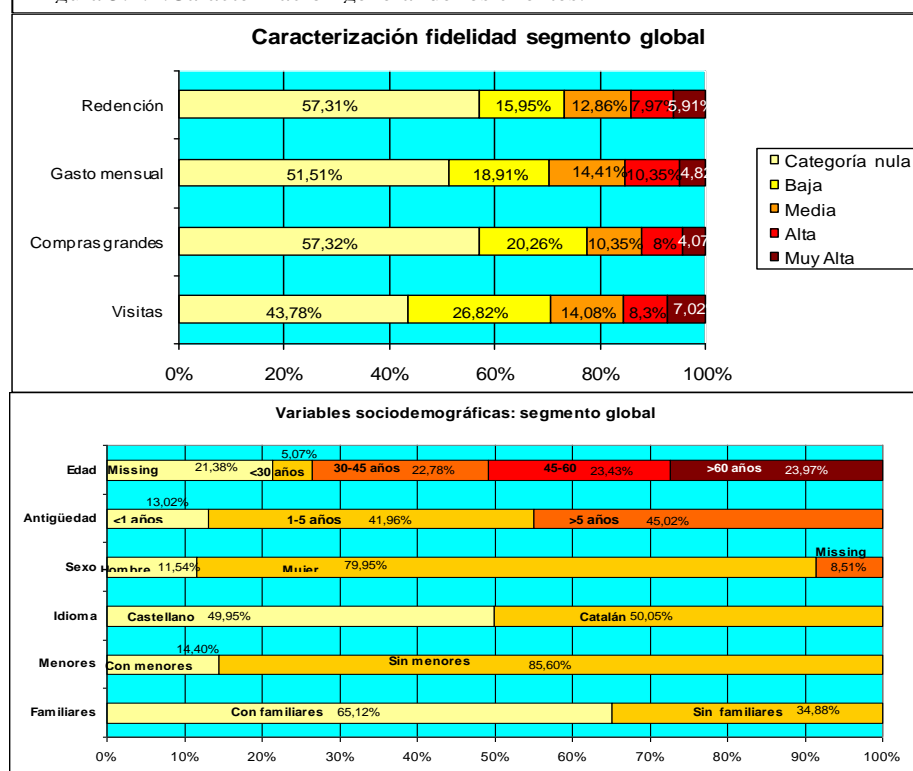
5. ESTUDIO DE LA REDENCIÓN

En este proyecto la variable redención adquiere gran importancia ya que esta variable nos marca el grado de sensibilidad al cupón de un cliente. En este apartado, se ha dividido a los clientes en 5 grupos dependiendo de su grado de sensibilidad al cupón y se ha caracterizado a cada grupo con variables de fidelidad y con variables sociodemográficas.

A continuación se caracterizará a cada grupo y se mostrará la distribución global de las variables de segmentación y sociodemográficas para todos los clientes con el objetivo de poder comparar la distribución general con la distribución de las variables de cada grupo. Además se presentará un resumen de los 6 grupos caracterizados.

5.1 CARACTERIZACIÓN GENERAL DE LOS CLIENTES

Figura 5.1.1. Caracterización general de los clientes.



En el primer gráfico se pone de manifiesto que la mitad de los clientes son poco fieles ya que gastan menos de 20€ al mes, casi no visitan el supermercado, no utilizan los cupones y no realizan ninguna compra grande en ninguna sección.

Respecto a la edad cabe destacar que existe un gran número de valores faltantes, que los menores de 30 años solo representan el 5% de toda la población y que el resto de grupos de edades se distribuyen de forma similar con un porcentaje de clientes del 23% aproximadamente.

Además se aprecia que el 45 % de todos los clientes tienen una antigüedad superior a 5 años y que solamente el 13% tiene una antigüedad inferior al año.

La mayoría de los clientes de la muestra son mujeres (79,9%) que viven en el hogar con familiares (65,12%) sin menores de edad (85,60%).

También es importante señalar que la mitad de los clientes utilizan el castellano como primera lengua y la otra mitad el catalán.

Figura 5.1.2. Resumen grupos en función de su nivel de redención.

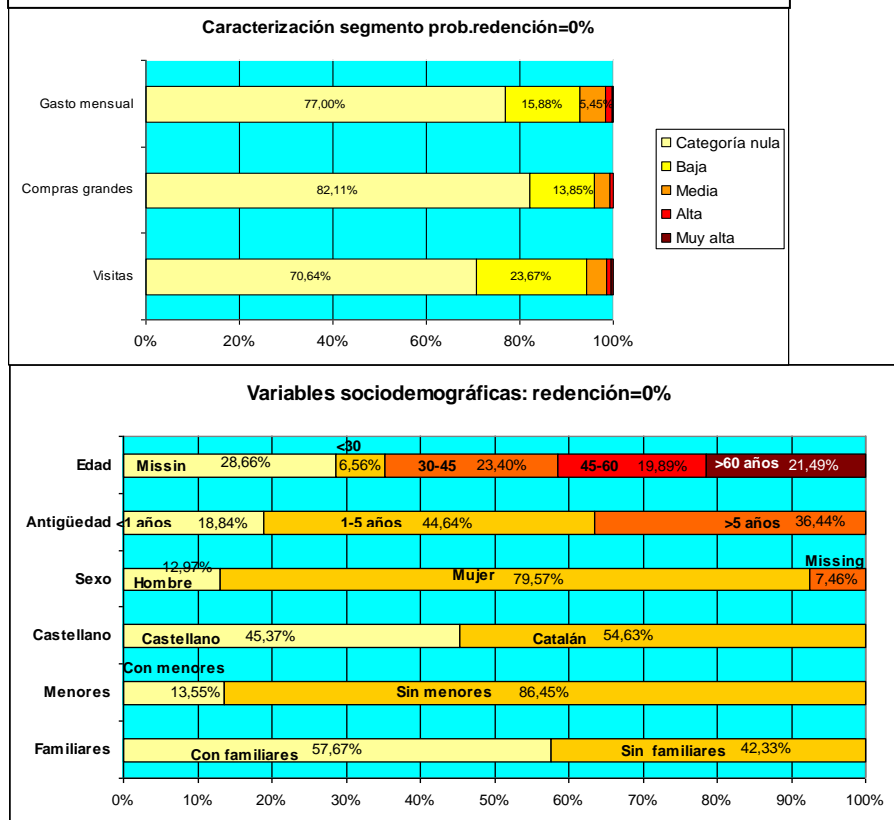
REDENCIÓN=0% (57.31%)	REDENCIÓN= (0%-3%] (15.95%)	REDENCIÓN= (3% -6%] (12,86%)
<ul style="list-style-type: none">-Grado de fidelidad muy reducido: gastan menos de 20€ al mes, realizan menos de una visita al mes y no efectúan ninguna compra grande en ninguna sección.-Utilizan el catalán como primera lengua.-Destaca el grupo de los menores de 30 años- Grupo con menor antigüedad.-Viven sin familiares.-Es el conglomerado que cuenta con el mayor porcentaje de hombres.	<ul style="list-style-type: none">-Gasto medio.-Es el segmento con el porcentaje mayor clientes que viven con menores de edad-Predominio de castellanoparlantes-Antigüedad mayor de 5 años.-Porcentaje elevado de clientes que viven con familiares en el hogar.	<ul style="list-style-type: none">-Es el grupo más fiel: el que más gasta y los que más compras grandes realizan.-Edad comprendida entre 45-60 años.-El porcentaje de menores de edad es similar a la media.-Predominio de castellanoparlantes-Antigüedad mayor de 5 años.-Porcentaje elevado de clientes que viven con familiares en el hogar.

REDENCIÓN= (6%-10%] (7.91%)	REDENCIÓN=(>=10%) (5.91%)
<ul style="list-style-type: none">-Grupo que más visitas realiza.-Gasto medio.-Gran porcentaje de clientes cuya edad es superior a 60 años.-Predominio de castellanoparlantes-Antigüedad mayor de 5 años.-Porcentaje elevado de clientes que viven con familiares en el hogar.	<ul style="list-style-type: none">-Son los segundos que más visitas realizan.-Gasto medio-bajo.-Grupo de mayor edad: grupo predominante de los mayores de 60 años.-Predominio de castellanoparlantes-Antigüedad mayor de 5 años.-Porcentaje elevado de clientes que viven con familiares en el hogar.

5.2 GRUPO SIN SENSIBILIDAD AL CUPÓN

Tamaño: 57.31%

Figura 5.2.1. Caracterización grupo sin sensibilidad al cupón.



Este grupo está formado por clientes con un **grado de fidelidad muy reducido**: gastan menos de 20€ al mes, no redimen, realizan menos de una visita al mes y no efectúan ninguna compra grande en ninguna sección.

Es el único grupo donde la mayoría de los clientes utilizan el **catalán como primera lengua**. También destaca un porcentaje superior a la media en el grupo de los menores de 30 años y tiene especial importancia el elevado porcentaje de valores faltantes, que no se distribuyen de forma aleatoria sino que se concentran especialmente en este grupo.

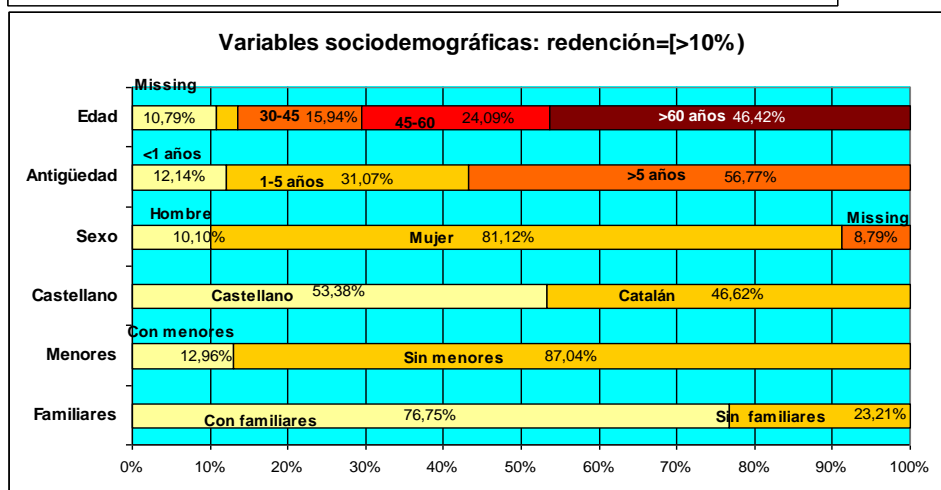
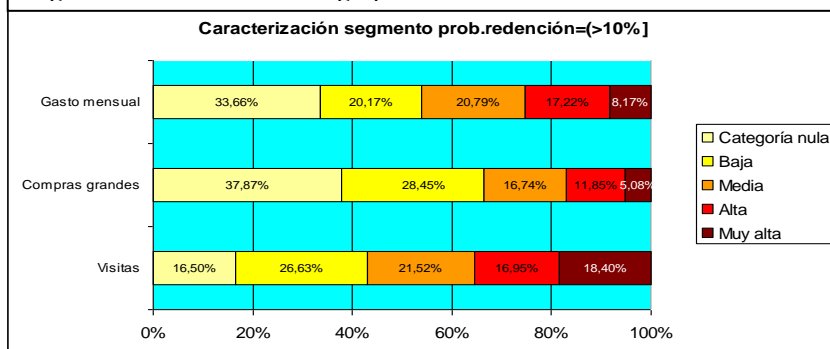
Es digno de mencionar que este es el grupo con menor antigüedad.

Además es el único grupo donde el **porcentaje de clientes que viven sin familiares** es elevado comparado con la media y también es el conglomerado que cuenta con el **mayor porcentaje de hombres**.

5.3 GRUPO CON LA MÁXIMA SENSIBILIDAD AL CUPÓN (Redención>10%)

Tamaño: 5.91%

Figura 5.3.1. Caracterización grupo redención>=10%.



Es el **grupo de mayor edad** de todos los considerados: el grupo predominante es el de los mayores de 60 años.

Se caracteriza además por ser los segundos que más visitas realizan y por tener un gasto medio-bajo.

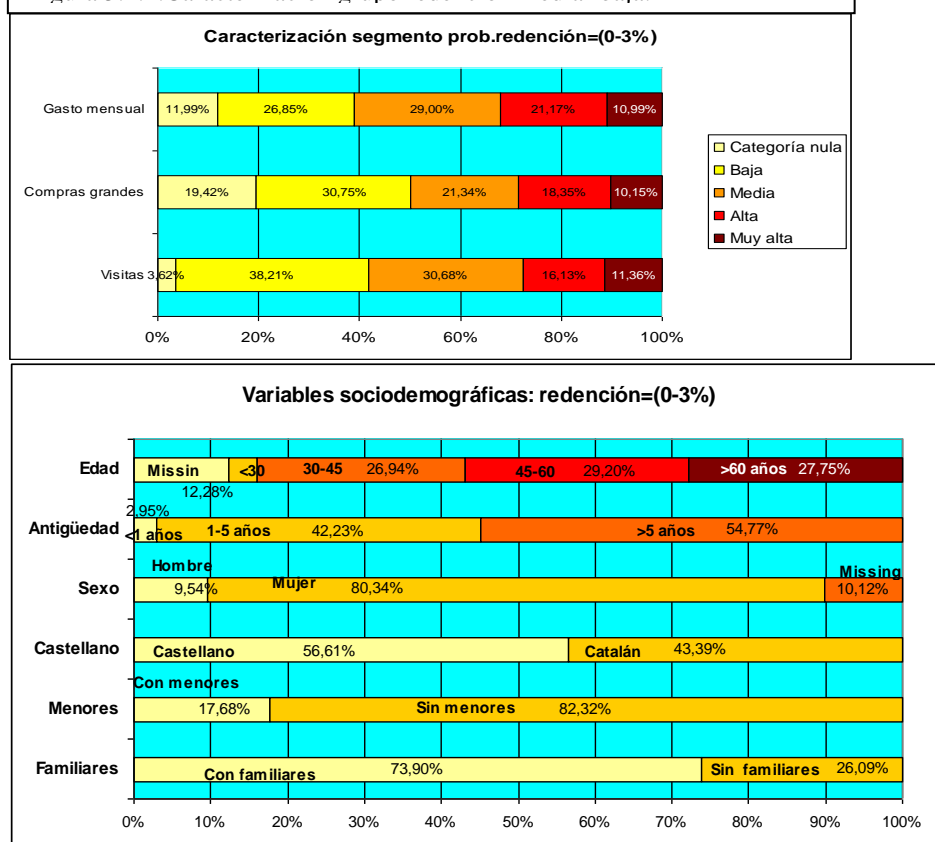
Al igual que los grupos que redimen que se presentan a continuación se caracterizan por lo siguiente:

- Predominio de castellanoparlantes
- Antigüedad mayor de 5 años.
- Porcentaje elevado de clientes que viven con familiares en el hogar.

5.4 GRUPO CON UNA SENSIBILIDAD AL CUPÓN REDUCIDA

Tamaño: 15.95%

Figura 5.4.1. Caracterización grupo redención media- baja.



Este grupo se caracteriza por tener un **gasto medio**, similar al grupo de redención alta (6%-10%].

Es un grupo con una edad menor comparado con los grupos que tienen una alta sensibilidad al cupón.

Es el segmento con el **porcentaje mayor clientes que viven con menores de edad** (17,65 frente a al 14 % de la media).

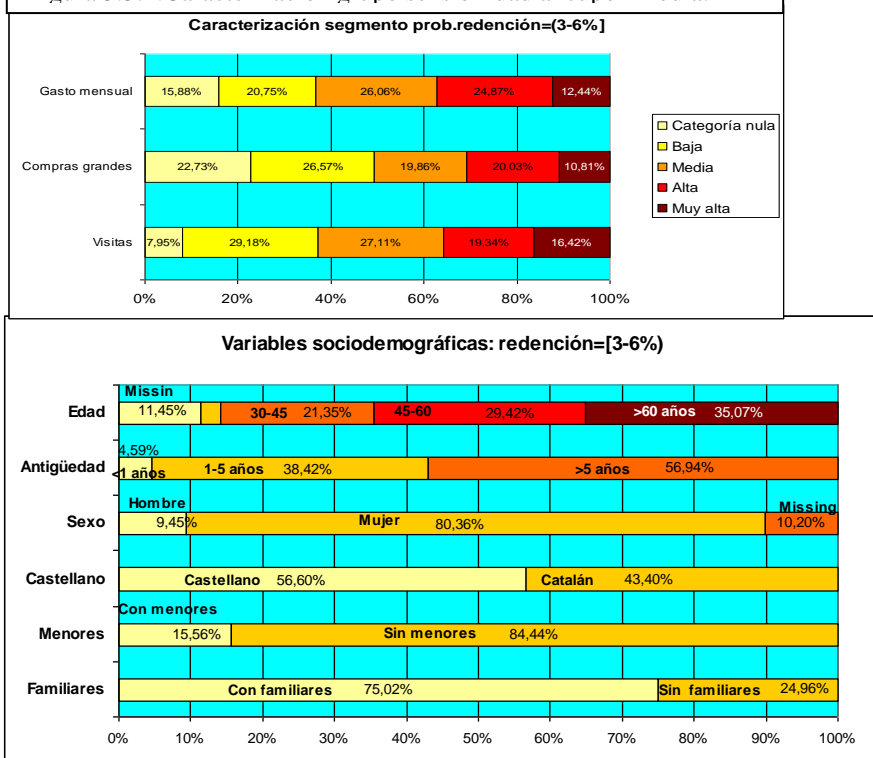
Al igual que los grupos que redimen que se presentan a continuación se caracterizan por lo siguiente:

- Predominio de castellanoparlantes
- Antigüedad mayor de 5 años.
- Porcentaje elevado de clientes que viven con familiares en el hogar.

5.5 GRUPO CON UNA SENSIBILIDAD AL CUPÓN MEDIA

Tamaño: 12.86%

Figura 5.5.1. Caracterización grupo sensibilidad al cupón media.



Este segmento es el **grupo más fiel**: el que más gasta y los que más compras grandes realizan.

Se caracteriza por tener el porcentaje más elevado de clientes con una **edad comprendida entre 45-60 años**.

El porcentaje de menores de edad es similar a la media.

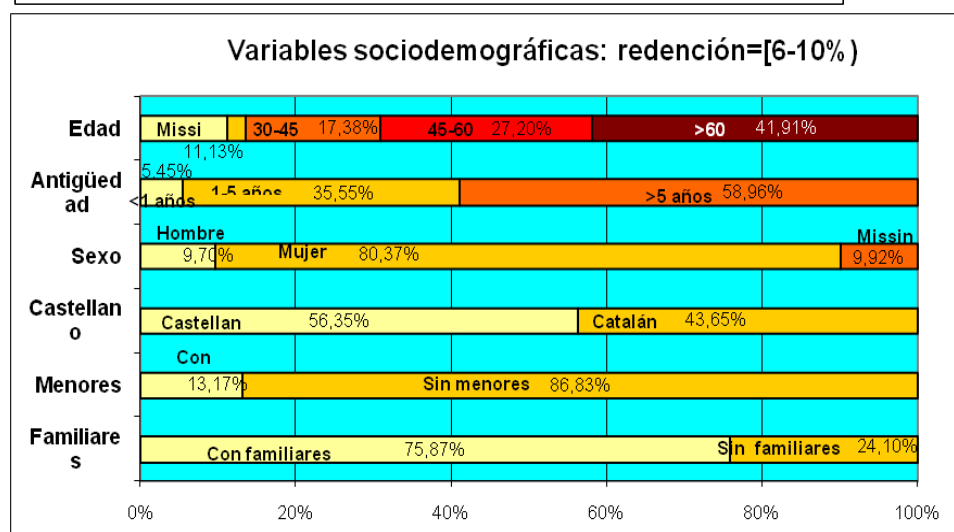
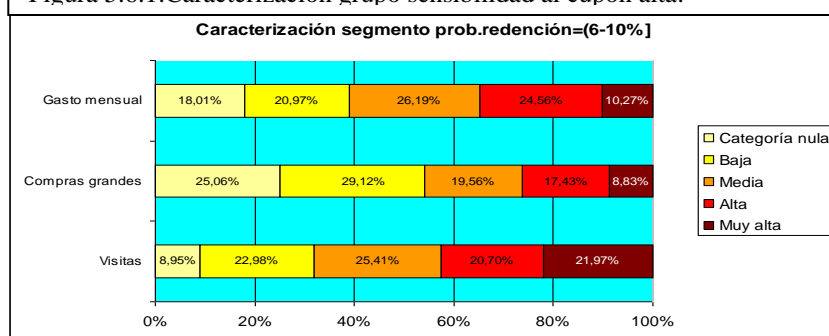
Al igual que los grupos que redimen que se presentan a continuación se caracterizan por lo siguiente:

- Predominio de castellanoparlantes
- Antigüedad mayor de 5 años.
- Porcentaje elevado de clientes que viven con familiares en el hogar.

5.6 GRUPO CON UNA SENSIBILIDAD AL CUPÓN ALTA

Tamaño: 7.97%

Figura 5.6.1. Caracterización grupo sensibilidad al cupón alta.



Es el grupo que más visitas realiza, caracterizado por tener un gasto medio similar al grupo de redención bajo y por tener un gran porcentaje de clientes cuya **edad es superior a 60 años**.

Al igual que los grupos que redimen que se presentan a continuación se caracterizan por lo siguiente:

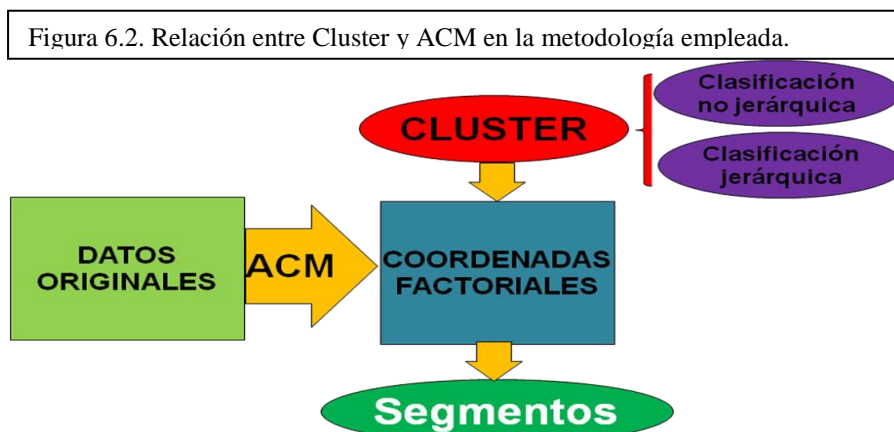
- Predominio de castellanoparlantes
- Antigüedad mayor de 5 años.
- Porcentaje elevado de clientes que viven con familiares en el hogar.

6.TÉCNICAS DE DATA MINING EMPLEADAS

En este capítulo se pretende hacer una breve revisión teórica sobre las principales técnicas de minería de datos que han sido empleadas en este trabajo. Básicamente como se aprecia en la figura 6.1 se han utilizado técnicas propiamente estadísticas como el Análisis Cluster y el Análisis de Correspondencias Múltiples y técnicas de Inteligencia Artificial como el Market Basket Analysis. En este capítulo se explicaran las dos primeras técnicas mientras que el Market Basket Analysis será explicado en el capítulo 10.



Como se observa en la figura 6.2 algo que será común en la mayoría de las segmentaciones que se presentan, es la aplicación del Cluster sobre las coordenadas factoriales procedentes de un ACM y no sobre las variables originales. En cada segmentación se explicará la metodología aplicada. Además como se desprende de la figura en este proyecto se han utilizado dos tipos de algoritmos de cluster: un algoritmo K-Means (Algoritmo de clasificación no jerárquica) y algoritmo de clasificación jerárquica según el criterio de Ward.



6.1 ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

Los análisis factoriales proporcionan una “rápida una fotografía privilegiada de la realidad multivariante que se está estudiando”. (Aluja, 1996) .De hecho los métodos factoriales nos dan información referente a las asociaciones entre variables proporcionándonos una gran aproximación visual destacando la información más relevante de la matriz de datos. (Aluja, 1996).

Dentro de estos métodos factoriales destaca el análisis de correspondencias múltiples (ACM). El ACM es una técnica descriptiva y multivariante que muestra un resumen de la información contenida en las tablas de contingencia cuando se dispone de un conjunto de I individuos caracterizados por J variables categóricas. (Escofier, et al. 1992)

Esta técnica de análisis de datos es muy útil cuando se trata de estudiar un conjunto de datos que contiene muchas variables categóricas, sobre todo para el análisis de encuestas.

La matriz que se presenta en la figura 6.1.1 es la llamada **codificación disyuntiva completa** o **matriz indicatriz binaria** y muestra la estructura de los datos antes de realizar un ACM.

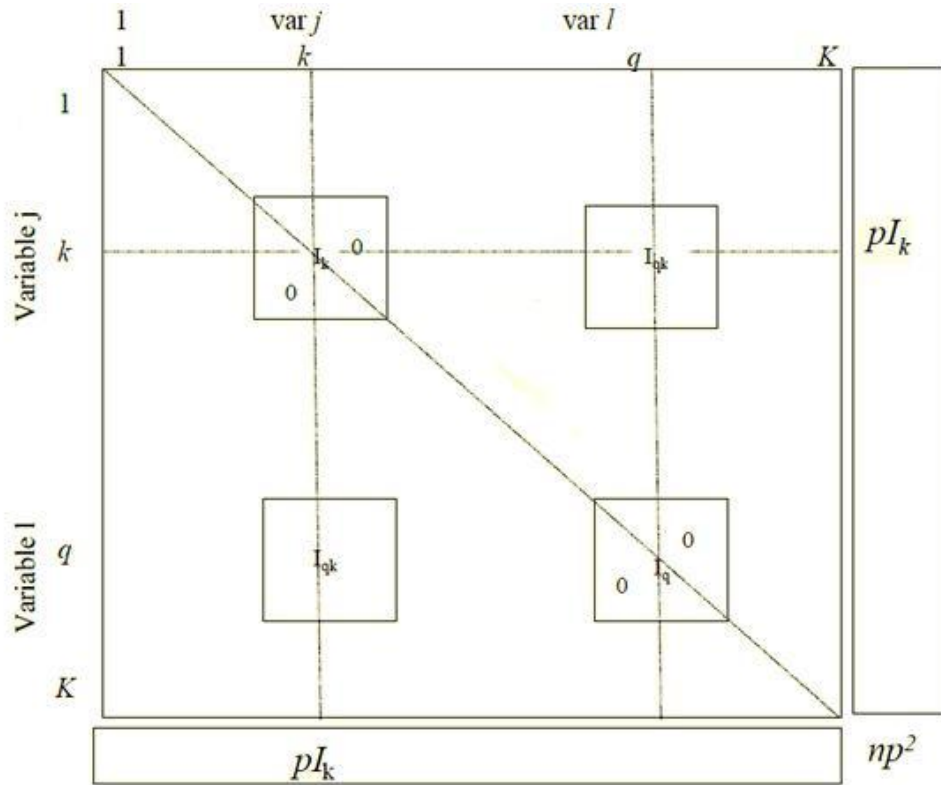
Para esta matriz existen tantas variables binarias (0/1) como modalidades tenga una variable categórica. Para un individuo determinado se codifica con un 1 si este cumple la modalidad correspondiente y con un 0 si no tiene esa modalidad.

Figura 6.1.1 Codificación disyuntiva completa.

	var 1	var 2	var k	Marge
Individuals				
1				p
i		001001	0001000	p
n				
Marge	l1	l2	lk	np

Como la tabla anterior es muy grande se puede generalizar y crear una tabla de dos dimensiones que contenga el conjunto de tablas de contingencia entre las variables tomadas 2 a 2. A continuación se muestra la tabla de Burt en las que las tablas que contienen la diagonal cruzan cada variable consigo misma. (Escofier, et al. 1992)

Figura 6.1.2 ACM. Tabla de Burt.



En un análisis factorial es clave analizar la nube de puntos fila y la nube de puntos columna.

La nube de puntos fila nos permite visualizar a los individuos en un plano factorial. Se dice que dos individuos serán parecidos si presentan globalmente las mismas características y si tienen posiciones cercanas en los ejes factoriales. La distancia entre dos individuos i y l viene definida por la siguiente expresión:

$$d^2(i, l) = \frac{1}{J} \sum_k \frac{I}{I_K} (x_{ik} - x_{lk})^2$$

La nube de puntos columna permite realizar la representación de las diferentes modalidades en los ejes factoriales. La distancia entre dos modalidades k y h se expresa de la siguiente forma:

$$d^2(k, h) = \sum_i I \left(\frac{x_{ik}}{I_K} - \frac{x_{ih}}{I_h} \right)^2$$

Además es interesante mencionar que cada individuo o modalidad influye en la creación de los planos factoriales a través de su inercia al centro de gravedad. Esta idea viene representada en la siguiente expresión:

$$\text{Inercia de } k \text{ con respecto a } G = \frac{1}{J} \left(1 - \frac{I_k}{I}\right)$$

Finalmente se exponen las relaciones de transición que permiten la representación simultánea de los puntos fila y columna y son básicos para la lectura de los ejes factoriales:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \frac{x_{ik}}{J} G_s(k)$$

Según la expresión anterior un individuo se sitúa en promedio en las coordenadas que asume, alejándose por la inversa de la raíz cuadrado del valor propio.

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \frac{x_{ik}}{I_k} F_s(i)$$

Según la otra relación anterior una modalidad se sitúa en promedio en las coordenadas de los puntos fila que toma, alejándose a través de la inversa de la raíz cuadrada de su valor propio. (*Escofier, et al. 1992*)

6.2 CLASIFICACIÓN AUTOMÁTICA

En el Data Mining la clasificación automática es una herramienta muy utilizada sobre todo para diferenciar individuos.

El potencial y la utilidad de esta técnica es que permite resumir la matriz de datos en diferentes grupos de individuos parecidos entre sí pero en el que cada grupo es diferente a otro. El clustering, por lo tanto, es un conjunto de procedimientos y algoritmos que se proponen clasificar o agrupar la matriz de datos en grupos o segmentos tales que:

- 1) Los individuos dentro de un conglomerado sean muy parecidos
- 2) Cada grupo sea relativamente diferente de otros.

Estos segmentos permiten una comprensión detallada de la estructura de los datos y a una mejor interpretación de los resultados.

En nuestro caso los clientes se agrupan bajo el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los diferentes grupos.

En este proyecto como medida de distancias se utiliza la distancia euclídea ya que es más fácil de interpretar.

Teniendo en cuenta las siguientes condiciones:

- 1) Separabilidad: $d(i, h) = 0 \Rightarrow ei = eh$
- 2) Simetría: $d(i, h) = d(h, i)$
- 3) Desigualdad triangular: $d(i, h) = d(i, e) + d(e, h)$
- 4) Condición de Krassner: $d(i, h) = \text{MAX}[d(i, e) + d(e, h)]$

La medida del grado de similitud entre las observaciones viene dada por:

- Índice de disimilaridad, en el caso en que se verifiquen las condiciones de separabilidad y simetría.
- Métrica o distancia, en el caso en que se verifiquen las condiciones de separabilidad, simetría y desigualdad triangular.
- Ultramétrica, en el caso en que se verifiquen las condiciones de separabilidad, simetría y la condición de Krassner.

Entre los criterios más utilizados para conseguir el índice de disimilaridad se encuentra la métrica de Minkowsky:

$$d_h(i, h) = \left(\sum_r |x_{ir} - x_{hr}|^\theta \right)^{\frac{1}{\theta}}$$

Con $\theta = 2$ distancia euclidia

Con $\theta = 1$ distancia city block

Con $\theta = \infty$ infinito distancia de Chebichev (Gherghi et al, 1999)

Además de lo explicado anteriormente conviene hablar de los **diferentes criterios de fusionar dos elementos**:

- 1) **Método de la unión simple** (distancia mínima): la distancia de c_1 de c_2 viene dada por la distancia más pequeña entre un elemento de c_1 y un elemento de c_2 ;
- 2) **Método de la unión completa** (distancia máxima): la distancia de c_1 de c_2 viene determinada por la distancia máxima entre un elemento de c_1 y un elemento de c_2 ;
- 3) **Método de la unión media**: la distancia de c_1 de c_2 viene establecida por la media de las distancias entre todos los puntos de c_1 y todos los puntos de c_2 .
- 4) **Método del centroide**: la distancia de c_1 y de c_2 es determinada por la distancia entre los baricentros de c_1 y c_2 ;
- 5) **Método Ward**: se basa en minimizar la variabilidad dentro de los grupos y maximizar la variabilidad entre los grupos, con el objetivo de conseguir grupos homogéneos en su interior y bien diferenciadas respecto a otros segmentos. (Gherghi et al, 1999)

6.2.1 ALGORITMOS DE CLASIFICACIÓN

Existe una gran variedad de algoritmos de clasificación pero los más empleados son los siguientes:

-Algoritmos de partición o métodos de clasificación no jerárquica: pueden determinar directamente las particiones de los individuos directamente. El más característico es el algoritmo K-Means.

-Algoritmo de clasificación jerárquica, se puede diferenciar en dos clases:

***Algoritmos de clasificación jerárquica ascendente:** procede a la construcción de segmentos por agregaciones de parejas de objetos

***Algoritmos clasificación jerárquica descendente:** construye los conglomerados por dicotomización del conjunto de los objetos.

En esta memoria explicaremos la clasificación ascendente jerárquica y los métodos de partición directa (algoritmo K-Means) ya que han sido los dos algoritmos utilizados para segmentar a los clientes.

6.2.1.1 MÉTODOS DE PARTICIÓN DIRECTA(K-MEANS)

El método de partición directa o clasificación no jerárquica es una técnica que permite obtener directamente una subdivisión en grupos de la nube de puntos. Existen diferentes métodos de clasificación no jerárquica entre los que destacan el **K-Means**.

El funcionamiento del algoritmo K-Means es el siguiente:

- 1) Se selecciona aleatoriamente los individuos que se considera como centros iniciales.
 - 2) Se calcula la distancia para cada individuo al centro asignándose el centro más cercano. En esta fase se define una primera partición.
 - 3) Para cada grupo formado en la etapa 2 se calcula su centro de gravedad que se considera como el nuevo centro y se repite la fase 2.
- El proceso sigue hasta la convergencia a una solución estable que se da cuando dos iteraciones definan la misma partición.

La distancia de un individuo i al centroide viene determinada por la distancia euclídea.

Este algoritmo es preferido en términos de rapidez computacional que los jerárquicos. No obstante, el fundamento de este algoritmo se basa en seleccionar de forma aleatoria los centros iniciales por lo que para aplicar esta método hay que tener una idea sobre el número de segmentos. (*Aluja, 1996*)

6.2.1.2 CLASIFICACIÓN ASCENDENTE JERÁRQUICA

Esta técnica consiste en agrupar los individuos que se quieren clasificar en grupos los cuales se van anidando en grupos posteriores, dando lugar a una forma de árbol. Se empieza el análisis con todos los individuos y posteriormente se van formando segmentos en sentido ascendente hasta que al llegar al final del proceso todos los casos tratados están encuadrados en un mismo grupo.

El algoritmo de clasificación ascendente jerárquica sigue el siguiente esquema:

- 1) Se calcula la matriz de distancias del conjunto de objetos que se quieren clasificar.
- 2) Se calcula la pareja de elementos del conjunto de datos que estén más próximos entre sí y se unen formando un nuevo bloque.
- 3) Se actualiza el conjunto formado y la matriz de distancias entre los elementos que existen y se vuelve a producir la unión de elementos bajo el criterio de la mínima distancia. De esta forma se origina un nuevo elemento en la jerarquía (nodo) y se va formando un árbol.

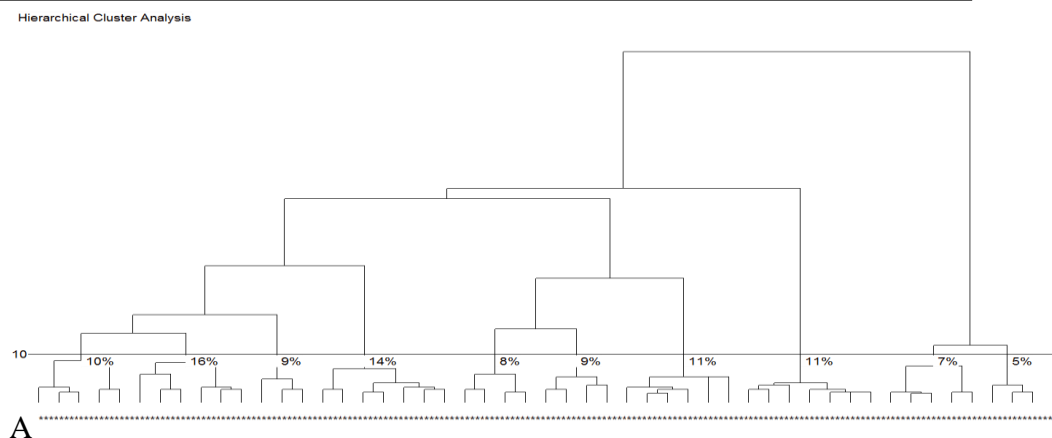
La representación gráfica de este proceso de acumulación de grupos se conoce con el nombre de dendograma que tiene forma de árbol invertido.

En los dendogramas adquiere gran importancia el índice de nivel que mide la distancia del nodo hasta los elementos que formaron dicho nodo. Índices de nivel altos sugieren que ha habido fusiones entre elementos muy diferentes entre sí e índices de nivel bajos indican señalan uniones de elementos semejantes.

El índice de nivel nos ayuda a localizar el punto donde efectuar el corte para identificar el número de clases a considerar.

Será conveniente efectuar el corte cuando existen saltos importantes en el índice de nivel. (Aluja, 1996)

Figura 6.2.1.2.1 Ejemplo de dendograma (cluster jerárquico).



7.SEGMENTACIÓN POR FIDELIDAD

7.1 DESCRIPCIÓN DE LA SEGMENTACIÓN

En este proyecto el primer objetivo planteado era diferenciar a los clientes según su grado de fidelidad distinguiendo aquellos segmentos de clientes que son sensibles a los cupones. Además la creación de segmentos ha de ser de tal forma que cada segmento tenga características comunes dentro del grupo al que pertenece y que presente disimilaridad con respecto a otros segmentos.

Aparte de esto, se caracterizará a los diferentes grupos con otras variables sociodemográficas y variables de fidelidad.

DATOS DE PARTIDA

Se parte de un dataset de compras efectuadas por clientes desde enero de 2009 hasta el 30 de septiembre de 2010.

- **Se excluyen a los clientes que tienen un gasto superior a 1.000€ y aquellos que su número de visitas mensuales es mayor que 10.**

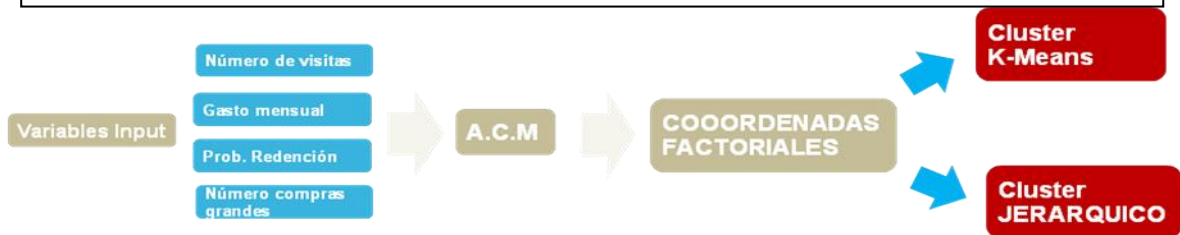
La explicación de esta restricción es la siguiente: no se está interesado en descubrir diferencias entre los que tienen un gasto muy elevado (mayor de 1,000€) y los que tienen un gasto más reducido o casi nulo (menos de 5 € mensuales). De esta forma nos centramos en las familias y excluimos a otras unidades de consumo como pueden ser los bares o restaurantes.

7.2 METODOLOGÍA

La metodología aplicada se puede resumir en los siguientes pasos:

- 1) Se parte de las variables de fidelidad: Gasto mensual, número de compras grandes, número de visitas y probabilidad de redención, tramificadas en 5 modalidades tal como se indica en el apartado 2.5.
- 2) Realización de un Análisis de Correspondencias Múltiples que nos permitirá pasar “de unas variables ordinales o categóricas observadas a un pequeño número de nuevas variables llamadas factores, tales que sinteticen la información de las variables originales” (Aluja, 1996).
- 3) Aplicación del algoritmo K-Means y cluster jerárquico sobre las coordenadas factoriales obtenidas de la aplicación del A.C.M y no sobre las variables originales.

Figura 7.2.1. Esquema metodología segmentación fidelidad.



Tal como se deduce de la figura presentada anteriormente se van a realizar dos segmentaciones utilizando dos algoritmos de cluster diferentes. Primero se probará con el K-Means y después se hará lo mismo con el algoritmo de clasificación ascendente jerárquico bajo el criterio de Ward.

La primera segmentación se realizó utilizando toda la muestra (188.275 clientes) en SAS ENTERPISE MINER con el procedimiento CLUSTERING NODE.

Como la aplicación del cluster jerárquico no es viable computacionalmente con 188.275 observaciones se ha seleccionado una muestra que contiene el 10% de todos los clientes (18.827 clientes) mediante un muestreo aleatorio simple y con esa muestra se ha realizado el cluster jerárquico después de realizar un Análisis de Correspondencias Múltiples. Esta segunda segmentación se ha realizado con el software de estadística multivariante SPAD.

7.3 RESULTADOS

7.3.1 RESULTADOS ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

El Análisis de Correspondencias Múltiple nos permite realizar una visualización gráfica y caracterización general en el espacio de las cuatro variables ordinales de fidelidad basada en la asociación entre las variables. Con este análisis podremos caracterizar a los clientes de acuerdo a su posicionamiento en los ejes factoriales. El Análisis de Correspondencias Múltiple ha sido realizado tanto en SAS como en SPAD. En este apartado se presentará el ACM realizado en SPAD. Los resultados en ambos son prácticamente idénticos.

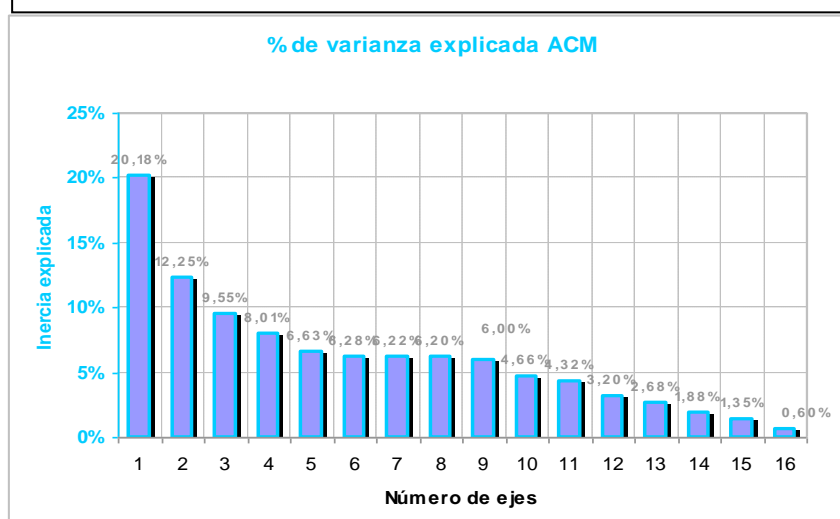
Para interpretar el Análisis de Correspondencias Múltiple se sigue el siguiente esquema:

- 1) Análisis de la descomposición de la inercia explicada por cada eje factorial.
- 2) Interpretación de los ejes factoriales mediante el análisis de las contribuciones de cada eje y proyectando las modalidades de las cuatro variables sobre los ejes factoriales.

TABLA DE DESCOMPOSICIÓN DE LA VARIANZA

Con el siguiente gráfico representamos la descomposición de la varianza explicada por cada eje. En él se puede observar que el primer eje factorial recoge gran parte de la inercia siendo el eje más diferenciado.

Figura 7.3.1.1 Gráfico porcentaje de varianza explicada.



Observamos que los porcentajes de inercia son bajos, debido a la existencia en el ACM de **“una gran parte de inercia trivial, debido a la asociación de cada variable consigo misma”** (Aluja, 1996). En el histograma se aprecia un primer eje muy diferenciado con respecto a los demás.

INTERPRETACIÓN DE LOS EJES FACTORIALES

ANÁLISIS DE LAS CONTRIBUCIONES

Con este análisis podremos observar qué modalidad contribuye más a la formación de los ejes factoriales:

Figura 7.3.1.2 Análisis de las contribuciones (resultados de SPAD).

Label	Relative Weight	Distance to origin	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Nivel de redención							
0%_red	14,328	0,7449	8,73	0,01	0	0,01	0,13
0%-3%_red	3,93	5,36132	4,78	1,02	0,25	0,01	24,38
3%-6%_red	3,313	6,54717	4,18	0,22	0	0,39	0,35
6%-10%_red	1,98	11,6263	2,38	0,27	0,68	0,11	13,53
>10%_red	1,45	16,2414	0,63	0	0,01	0	34,21
			20,69	1,53	0,95	0,53	72,59
Visitas mensuales							
<1_visita	11,008	1,27118	11,15	2,77	0,67	0,2	0,03
1-3_visita	6,6	2,78788	0,11	9,18	3,29	1,15	0,1
3-5_visita	3,555	6,03235	4,68	1,78	1,22	0,24	9,96
5-7_visita	2,048	11,21	4,75	0,81	0,53	0,1	0,9
>7_visita	1,79	12,9665	5,18	6,79	0,78	0,33	10,41
			25,87	21,33	6,5	2,02	21,4
Compras grandes							
0_GC	14,243	0,75531	10	1,08	0,08	0	0,05
1-3_GC	5,058	3,94315	1,81	16,98	6,68	0,85	1,73
3-6_GC	2,675	8,34579	4,75	0,66	12,88	19,57	0,06
6-9_GC	2,023	11,3609	5,83	4,32	5,47	23,8	1,43
>9_GC	1,003	23,9377	3,64	13,73	19,89	3,91	0,63
			26,04	36,78	45,01	48,14	3,91
Gasto mensual							
<20€	12,905	0,93723	11,11	2,15	0,47	0,11	0,02
20€-50€	4,55	4,49451	0,33	14,88	12,18	6,81	0,72
50€-100€	3,76	5,64894	4,87	5,19	5,92	21,75	0,01
100€-200€	2,56	8,76563	6,93	4,06	10,07	17,38	0,01
200€-1000€	1,225	19,4082	4,16	14,09	18,91	3,26	1,34
TOTAL			27,39	40,36	47,54	49,31	2,09

Las modalidades que más contribuyen al eje 1 son:

- No realizan compras grandes.
- Gasto mensual inferior a 20€.
- No redimen.
- Visitan menos de una vez al mes el supermercado.

EJE1	
-	+
<1 visitas	3-6 GC
<20€	6-9 GC
Red=0%	3-5 visitas
0_GC	5-7 visitas

Las modalidades que más contribuyen al eje 2 son:

- Realizan entre 1 y 3 compras grandes.
- Realizan entre 1 y 3 visitas.
- Gasto mensual entre 20-50€.
- Gasto mensual entre 200€-1000€.
- Realizan más de 7 visitas.

EJE2	
-	+
1-3 GC	6-9 GC
1-3 visitas	7-10 visitas
20€-50€	200€-1.000€

Las modalidades que más contribuyen al eje 3 son:

- Gasto mensual entre 200€-1.000€.
- Gasto mensual entre 20€-50€.
- Realizan más de 9 compras grandes.
- Realizan entre 3 y 6 compras grandes.

EJE3	
-	+
100€-200€	200€-1.000€
3-6 GC	>9 GC

Las modalidades que más contribuyen al eje 4 son:

- Realizan entre 6 y 9 compras grandes.
- Gasto mensual entre 100€-200€.
- Realizan entre 3 y 6 compras grandes.
- Gasto mensual entre 50-100€.

EJE4	
-	+
100€-200€	50€-100€
6-9 GC	3-6 GC

Las modalidades que más contribuyen al eje 5 son:

- Redención 0-3%
- Redención >10%

EJE5	
-	+
Red=(0-3 %)	Red>10%

Como hemos comentado anteriormente el ACM permite analizar las asociaciones entre individuos y modalidades de variables mediante la representación gráfica proyectando en los ejes factoriales los individuos y las variables.

En el gráfico 7.3.1.3 que se expone a continuación se proyectan las diferentes modalidades de cada variable de fidelidad y los individuos en dos ejes factoriales.

Se aprecia claramente en la figura 7.3.1.3 que las modalidades más bajas de las variables de fidelidad (0 % de redención, gasto mensual inferior a 20€, menos de una visita al mes y número de compras grandes inferior a 1) se sitúan a la izquierda del eje1. En la derecha por el contrario se sitúan las modalidades con un nivel de fidelidad más elevado.

El eje1 contrapone clientes cuyo grado de fidelidad es nulo (gasto inferior a 20€, redención nula, sin compras grandes en ninguna sección y con menos de una visita al mes) **contra clientes cuyo grado de fidelidad no es nulo** (realizan más de 1 visita al mes, gastan más de 20€ al mes, han redimido alguna vez y han realizado alguna compra grande en una sección).

El eje 2 excluye aquellos clientes cuyo grado de fidelidad es nulo. Además muestra oposición entre clientes que tienen un gasto bajo, visitas reducidas y compras considerables reducidas contra clientes que tienen un gasto elevado, visitan más de 7 veces al mes y realizan entre 6 y 9 compras grandes.

Es un eje que nos **marca la fidelidad excluyendo aquellos clientes con un grado de fidelidad nulo y nos opone clientes cuyo grado de fidelidad es bajo contra aquellos clientes cuyo grado de fidelidad es alto.**

Observando los gráficos se aprecia que existe un claro **efecto Gutman**. El efecto Gutman viene determinado por una **presentación en forma parabólica de los puntos representados**, es decir, en la parte izquierda (elipse A) y en la derecha (elipse C) aparecen las modalidades extremas y en la parte inferior aparecen las modalidades intermedias (elipse B). Con esto se puede decir que la estructura de los datos **es casi unidimensional** y existe una dominación del efecto tamaño ya que el eje 1 es el que más explica la variabilidad de los datos. Este efecto Gutman hace que la interpretación del eje 2 no esté muy clara.

Las variables gasto mensual, número de visitas y número de compras grandes muestran una estructura jerárquica y siguen claramente ese efecto Gutman. En cambio la variable redención se escapa de ese efecto mostrando una estructura diferente que se puede apreciar en el gráfico 7.3.1.3 en la que se representa la variable redención con un color amarillo.

Lo más evidente es que aquellos clientes que no redimen se caracterizan por un nivel de redención nulo. Parece que los que tienen un porcentaje de redención medio (3-6% de redención) tienen un grado de fidelidad mayor que aquellos que redimen más del 10%. Esto significa que aquellos clientes caracterizados por ser sensibles a los cupones (redención mayor del 10%) tienden a tener un gasto inferior con respecto a los que tienen una sensibilidad media a los cupones.

En la figura 7.3.1.4 se amplía el zoom para poder observar con mayor detalle las relaciones de asociación entre las variables suplementarias.

En general se aprecia que todas las modalidades se concentran en el centro de la nube de puntos, por lo que se ha tenido que ampliar el zoom para poder observar las relaciones de asociación entre las variables.

A la vista de este gráfico se deduce que el hecho de hablar castellano como primera lengua supone un grado de fidelidad más elevado que el hecho de tener el catalán como primera lengua, ya que la modalidad Catalán está más próxima de las modalidades que indican ausencia de fidelidad (gasto inferior a 20€, redención nula, número de compras grandes inferior a uno y número de visitas mensuales inferior a uno).

Una antigüedad superior a 5 años implica un grado de fidelidad mayor con respecto a los clientes que solo llevan un año con la tarjeta de fidelidad.

Respecto a la edad se aprecia que los jóvenes es el grupo de edad menos fiel y los clientes de 45-60 años son aquellos cuyo grado de fidelidad es mayor. Además el hecho de vivir con familiares y con menores de edad en el hogar implica un mayor grado de fidelidad respecto a vivir sin menores y sin familiares.

En los gráficos 7.3.1.5 y 7.3.1.6 se presenta la proyección de las variables activas sobre el eje1 y el 3 y también sobre el eje 1 y el eje5. Como se cita a continuación el eje 5 se interpreta como el nivel de redención y el eje 3 opone clientes que compran todo en el supermercado contra aquellos que son más selectivos y no realizan todas sus compras en el supermercado.

Figura 7.3.1.3: Proyección de las variables activas y los individuos.

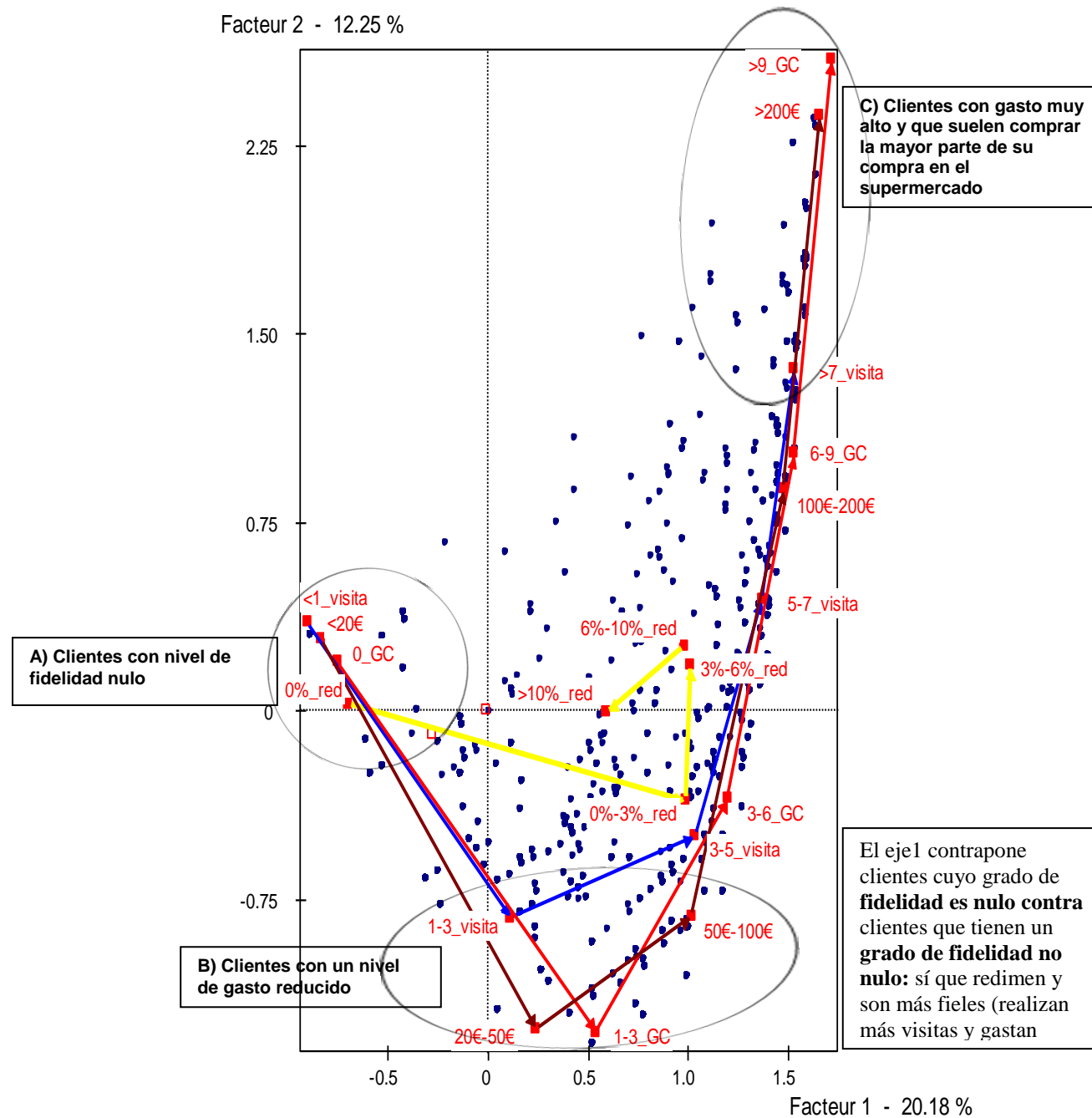


Figura 7.3.1.4: Proyección de las variables suplementarias sociodemográficas.

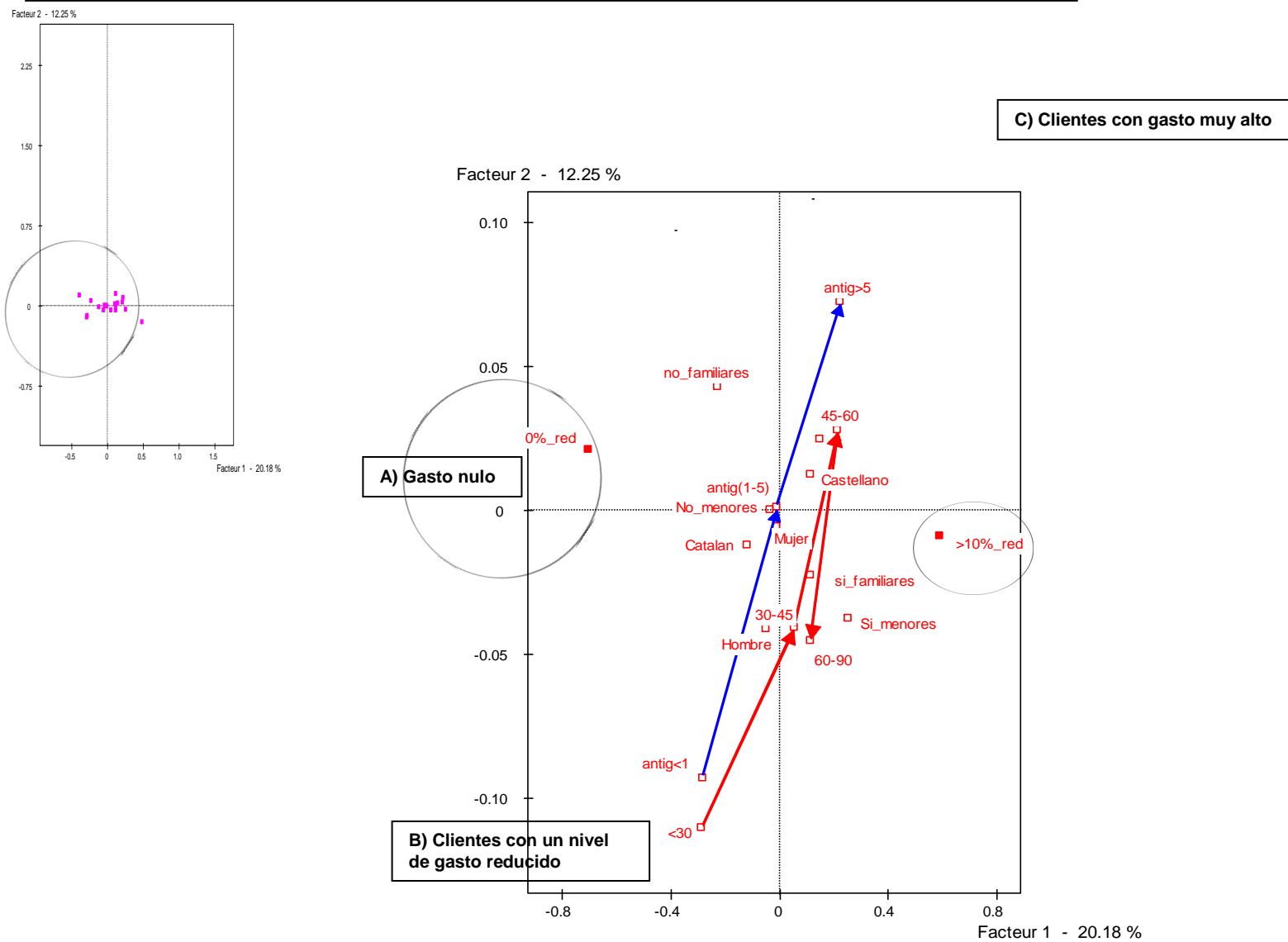


Figura 7.3.1.5. Proyección de las variables activas sobre el eje 1 y 3.

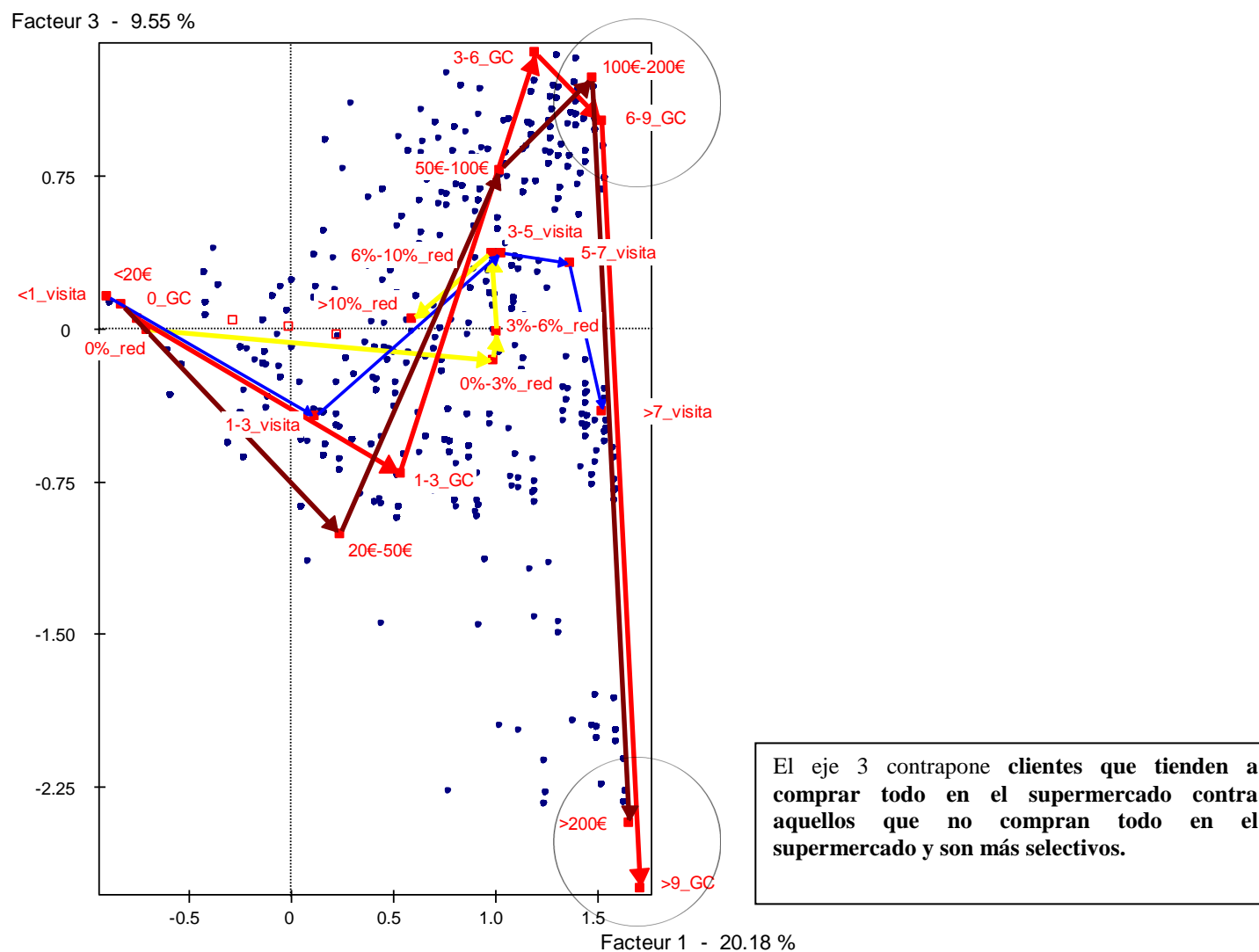
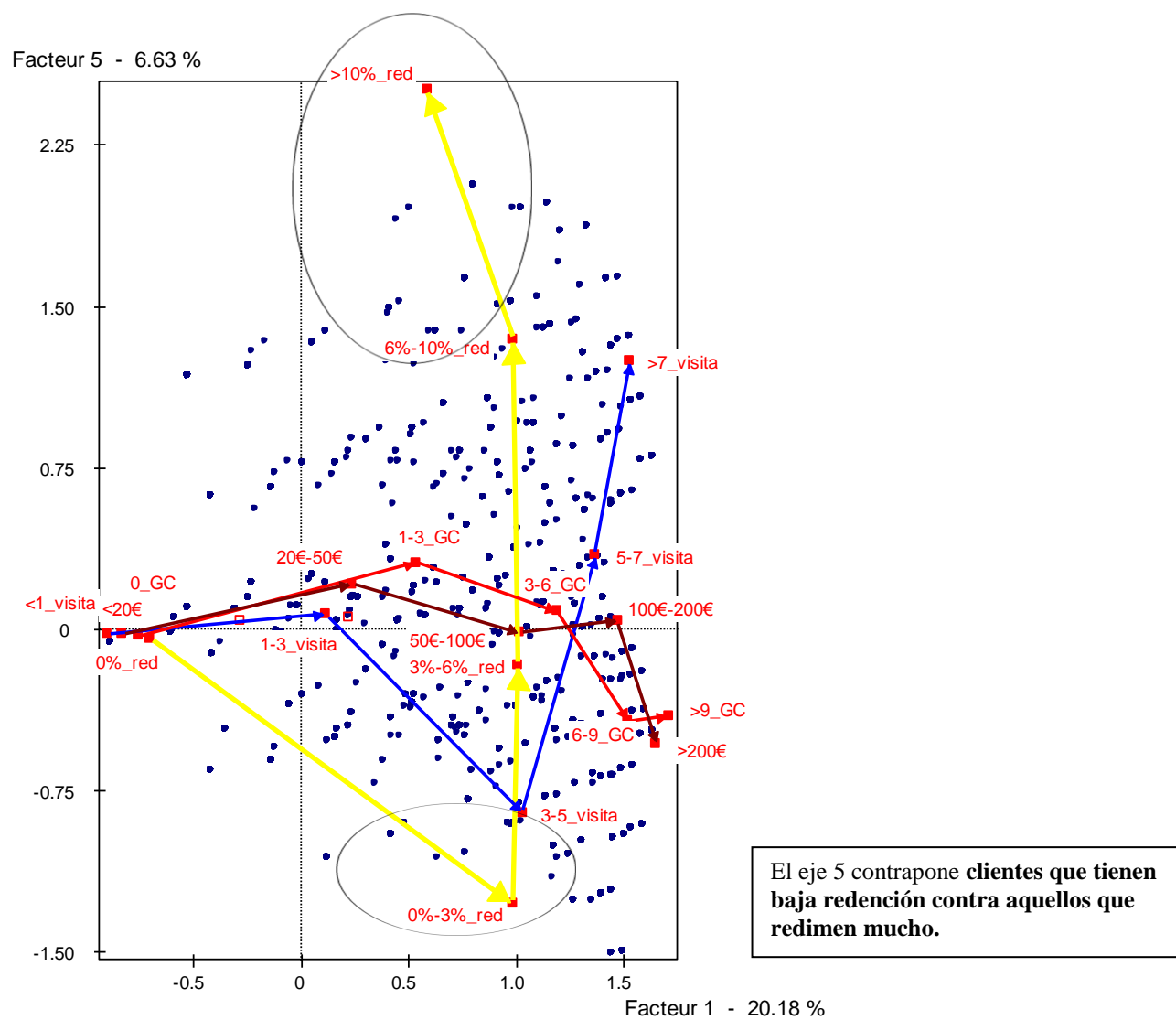


Figura 7.3.1.6. Proyección de las variables activas sobre el eje 1 y el eje5.



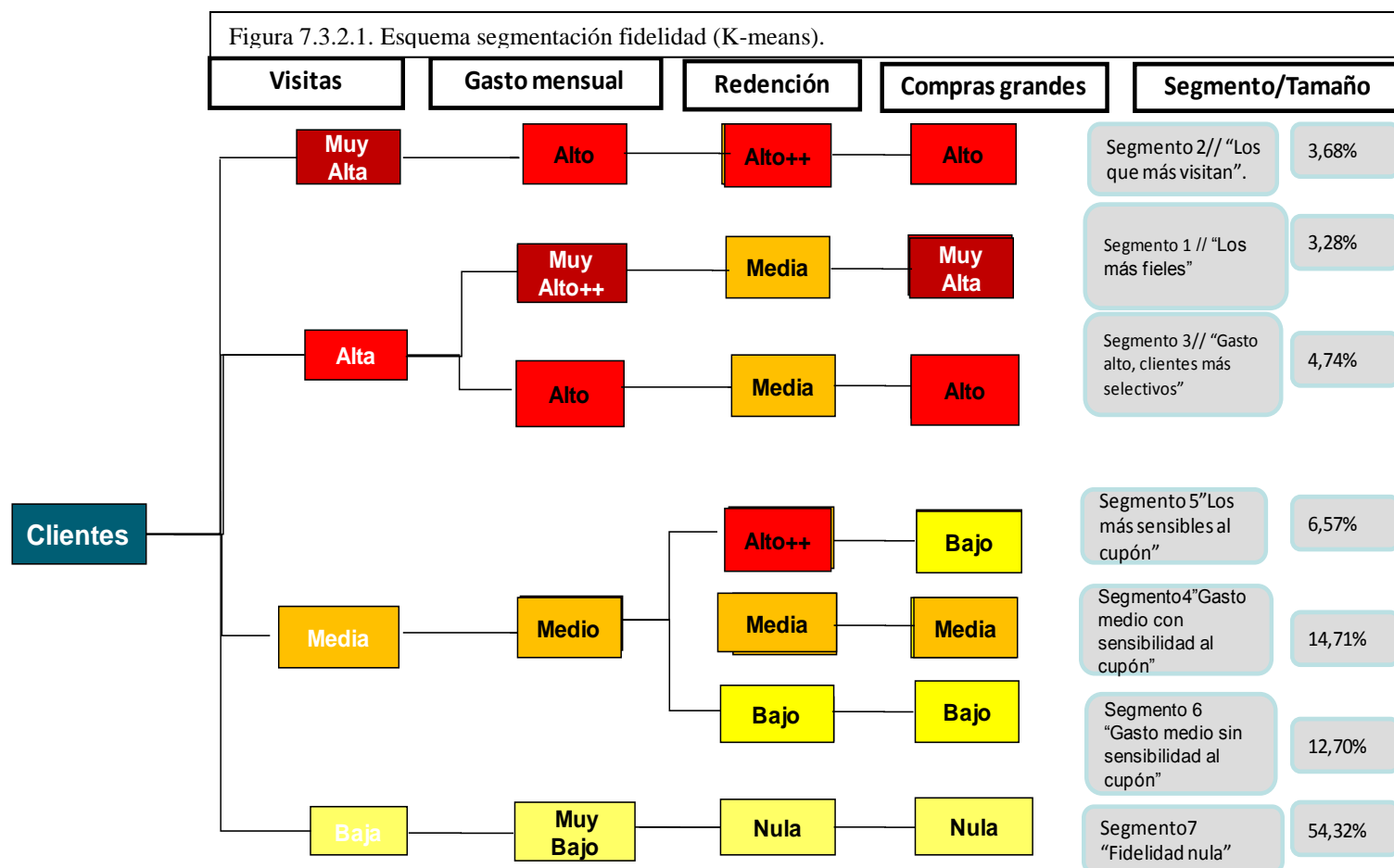
7.3.2 RESULTADOS SEGMENTACIÓN CON ALGORITMO K-MEANS

Después de haber realizado el Análisis de Correspondencias Múltiples a través del procedimiento `proc corresp` de SAS, se aplica el algoritmo K-Means a todos los individuos **considerando los primeros 10 ejes factoriales** del ACM. Para ello se ha utilizado el nodo clustering de SAS MINER.

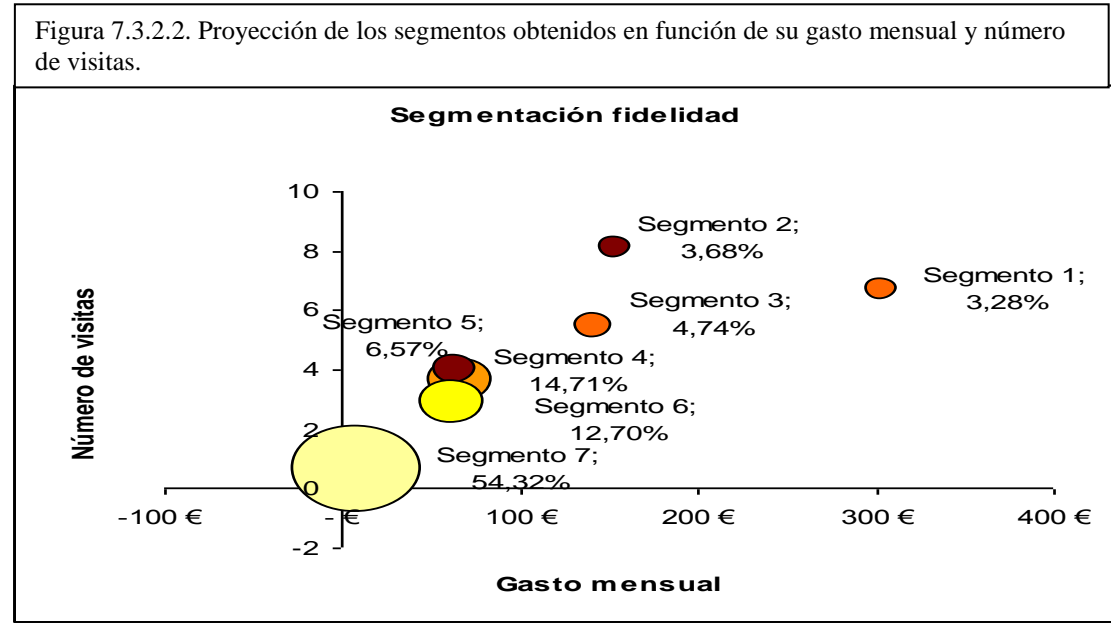
La siguiente figura que se adjunta a continuación muestra de forma esquemática un resumen de los segmentos obtenidos. Así observando dicho esquema, se aprecia que existe un segmento (segmento número 2) que son los que más visitas realizan, tienen un nivel de redención alto y un gasto elevado (100€-200€). Además destaca el segmento 1 por tener el nivel de fidelidad más alto: los que más gastan y los que realizan casi todas sus compras en el supermercado pero sin embargo no tienen un nivel de visitas tan elevado como el segmento 2. El segmento 3 tiene un grado de fidelidad alto en general pero es un segmento más selectivo en el sentido de que no compran absolutamente todo en el supermercado, aunque tiene un gasto elevado.

Además existen tres segmentos que se caracterizan por tener un número de visitas y un gasto mensual medio pero que difieren en el nivel de redención. Así pues, tenemos el segmento 5 caracterizado por un grado de fidelidad alto, el segmento 4 con un nivel de redención medio y el segmento 6 con un grado de sensibilidad al cupón débil.

También hay que señalar la existencia del segmento 7 que es el más numeroso y el menos fiel.



El gráfico 7.3.2.2 representa el gasto mensual medio y el número de visitas medio de cada segmento. Además el tamaño de cada segmento es proporcional al porcentaje de clientes que forma parte de cada conglomerado. También el color de cada segmento nos indica el grado de redención, es decir, el color rojo oscuro nos indica que es un segmento muy sensible a los cupones mientras que un color amarillo nos indica que es un grupo caracterizado por una escasa sensibilidad al cupón.



En esta tabla se proporciona un resumen descriptivo con el porcentaje de clientes que pertenece a cada segmento, el promedio del número de visitas, el promedio del nivel de redención y la media del número de compras grandes.

Figura 7.3.2.3. Tabla resumen de cada segmento.

Segmentos	% clientes	Visitas	Gasto mensual	Redención	Compras Grandes	Etiquetas
Segmento 1	3,28%	6,71	304,27 €	4,44%	10,5	Los más fieles".
Segmento 2	3,68%	8,11	153,55 €	7,44%	6,8	Los que más visitan: muchas visitas para tan poco gasto"
Segmento 3	4,74%	5,50	141,04 €	5,19%	6,7	Gasto elevado , clientes más selectivos"
Segmento 4	14,71%	3,63	66,76 €	3,26%	3,2	Gasto medio:redención media"
Segmento 5	6,57%	4,01	63,53 €	7,58%	2,6	Los que más redimen"
Segmento 6	12,70%	2,88	61,73 €	2,03%	2,7	Gasto medio:redención baja"
Segmento 7	54,32%	0,61	7,99 €	1,01%	0,1	Los menos fieles"

En la figura 7.3.2.4 se presenta un resumen con las características principales de los segmentos obtenidos. Además se proporciona a partir del apartado 7.3.2.1 información más detallada de cada segmento.

Figura 7.3.2.4. Resumen segmentos obtenidos

GASTO ALTO

SEGMENTO 1 (3,28%)

- Realiza casi todas sus compras en el supermercado.
- Gasto muy alto: superior a 200€.
- Visitas altas.
- Gasto por visita muy alto: 50€ de promedio.
- Redención media.
- Edad 45-60 años
- Antigüedad>5 años.
- Castellanoparlantes.
- Viven con menores en el hogar.
- Viven con familiares en el hogar.

SEGMENTO 2 (3,68%)

- Los que más visitan, más de 7 visitas al mes de promedio.
- Gasto alto y sensibilidad alta al cupón.
- Gasto por visita reducido.
- Edad superior a 60 años.
- Antigüedad>5 años.
- Castellanoparlantes.
- Viven con menores y con familiares en el hogar.

SEGMENTO 3 (4,74%)

- Gasto alto pero no compran en todas las secciones.
- Nivel de redención medio.
- Predominio del grupo de edad superior a 60 años.
- Son más jóvenes que el segmento 2.
- Antigüedad>5 años.
- Castellanoparlantes.
- Viven con menores y con familiares en el hogar.

GASTO MEDIO

SEGMENTO 4 (14,71%)

- Gasto medio con sensibilidad media a los cupones.
- Edad similar a la distribución global.
- Porcentaje de hombres superior a la media.
- Predominio de castellanoparlantes.

SEGMENTO 5 (6,57%)

- Gasto medio y nivel muy alto de sensibilidad a los cupones.
- El grupo de mayor edad.
- Predominio de mayores de 60 años.
- Viven sin menores de edad.
- Predominio de castellanoparlantes.
- Viven con familiares.
- Antigüedad más elevada que el segmento 4 y 6.

SEGMENTO 6 (12,70%)

- Gasto medio y escasa sensibilidad a los cupones.
- Más jóvenes que los grupos 5 y 4.
- Porcentaje de menores superior que el segmento 5.

GASTO REDUCIDO

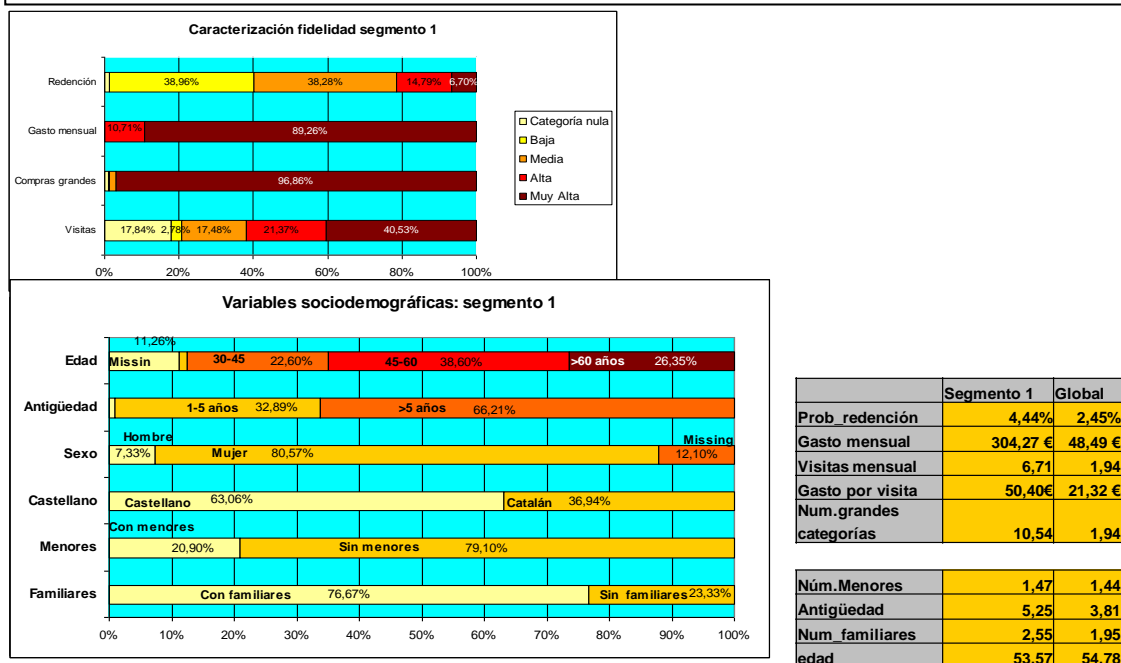
SEGMENTO 7 (54,32%)

- Segmento menos fiel.
- Segmento más numeroso: el 54,32%.
- Predominio de catalanoparlantes.
- Viven sin familiares.
- Viven sin menores de edad.
- Grupo con menor antigüedad con la tarjeta de fidelidad.

7.3.2.1 DESCRIPCIÓN SEGMENTO 1: “Los más fieles”

Tamaño: 3.28%

Figura 7.3.2.1. Caracterización segmento 1



Es el segmento que realiza casi todas sus compras en el supermercado y que se caracteriza por tener un gasto superior a 200€.

Tiene también un número de visitas altas (promedio de 6,7 visitas) aunque no es el grupo que más visita.

En cuanto a la redención se puede decir que en general se caracterizan por tener un nivel de redención medio.

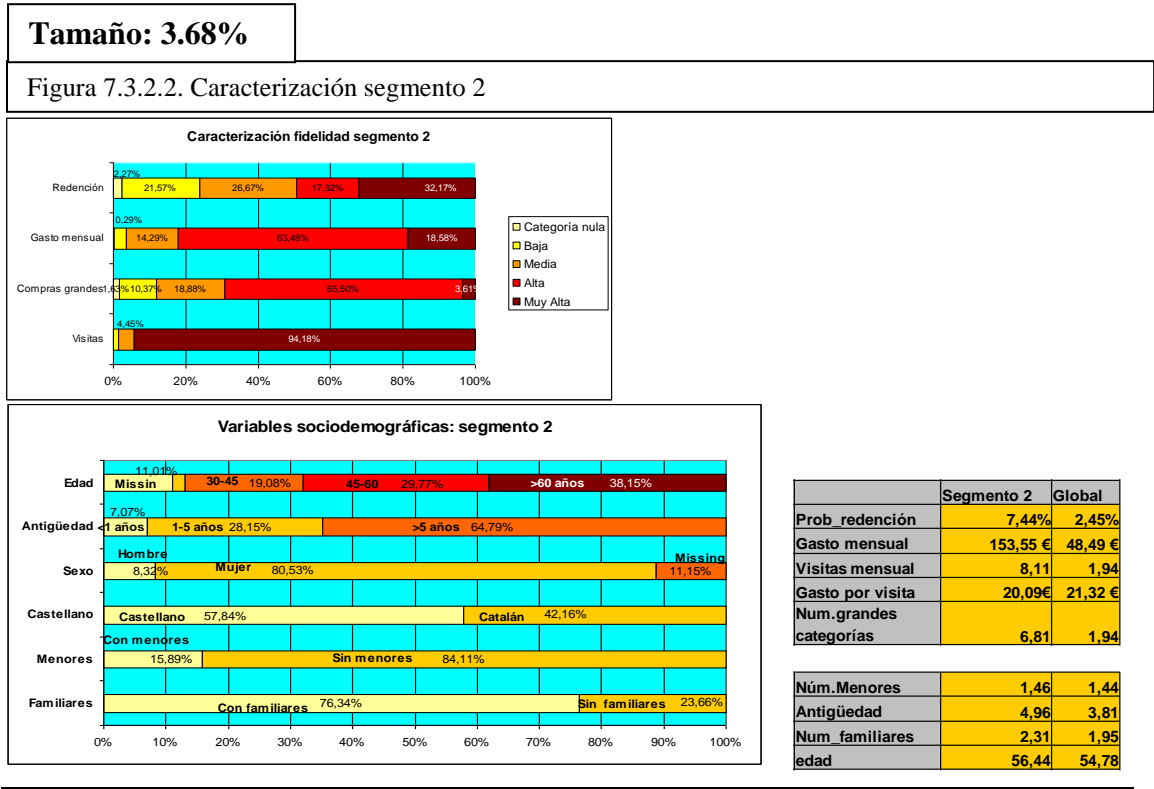
Respecto a la edad existe un predominio en el grupo de edad de 45-60 años y además el porcentaje de menores de 30 años es nulo.

En lo que a la antigüedad se refiere se caracteriza por tener una antigüedad superior a 5 años. Es el grupo con mayor antigüedad.

Un hecho que resulta curioso es el predominio de un porcentaje considerable de castellanoparlantes (63% de castellanoparlantes). Es el grupo que cuenta con el mayor porcentaje de castellanohablantes.

Además se aprecia que este segmento tiene el mayor porcentaje de clientes que viven con menores de hogar y que viven con familiares.

7.3.2.2 DESCRIPCIÓN SEGMENTO 2 “Los que más visitan”



Es el grupo que más visitas realiza al mes en promedio.

Además son clientes selectivos que, aunque tienen un gasto alto, no compran todo en el supermercado.

En general son bastantes sensibles al cupón: el 50% de todos los clientes tiene un nivel de redención alta.

Respecto a la edad destaca el grupo con una edad superior a 60 años. Es el segundo segmento con mayor edad.

Respecto a la antigüedad se caracteriza por tener una antigüedad superior a 5 años.

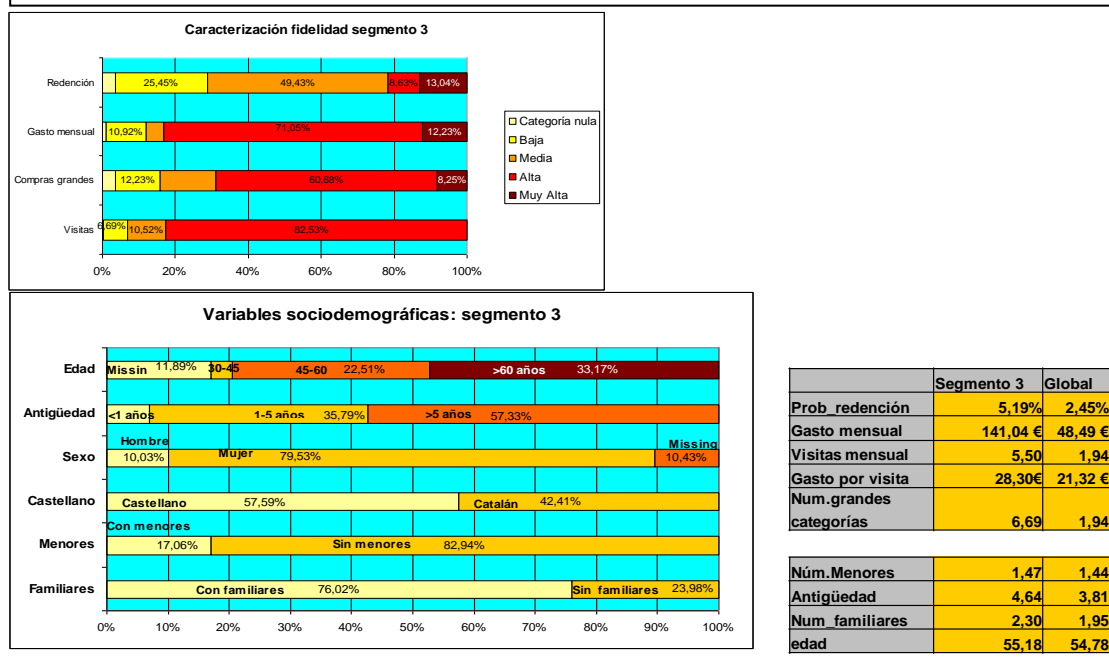
Además sigue destacando el predominio de los castellanohablantes.

El porcentaje de clientes con menores de edad es similar a la distribución media y como ocurría en el grupo anterior el porcentaje de clientes que viven con familiares es superior a la media.

7.3.2.3 DESCRIPCIÓN SEGMENTO 3 : “Gasto alto, clientes más selectivos”

Tamaño: 4.79%

Figura 7.3.2.3. Caracterización segmento 3



Este grupo se caracteriza por tener un gasto mensual alto (100-200€ al mes). Son clientes que no tienden a comprar todo en el supermercado.

Se caracterizan por tener un nivel de redención medio, no tan alto como el segmento 2 y por un nivel de visitas medio-alto. Además no son tan visitantes como el grupo 2.

Respecto a la edad se aprecia que son más jóvenes que el grupo 2, aunque existe mayoría de mayores de 60 años.

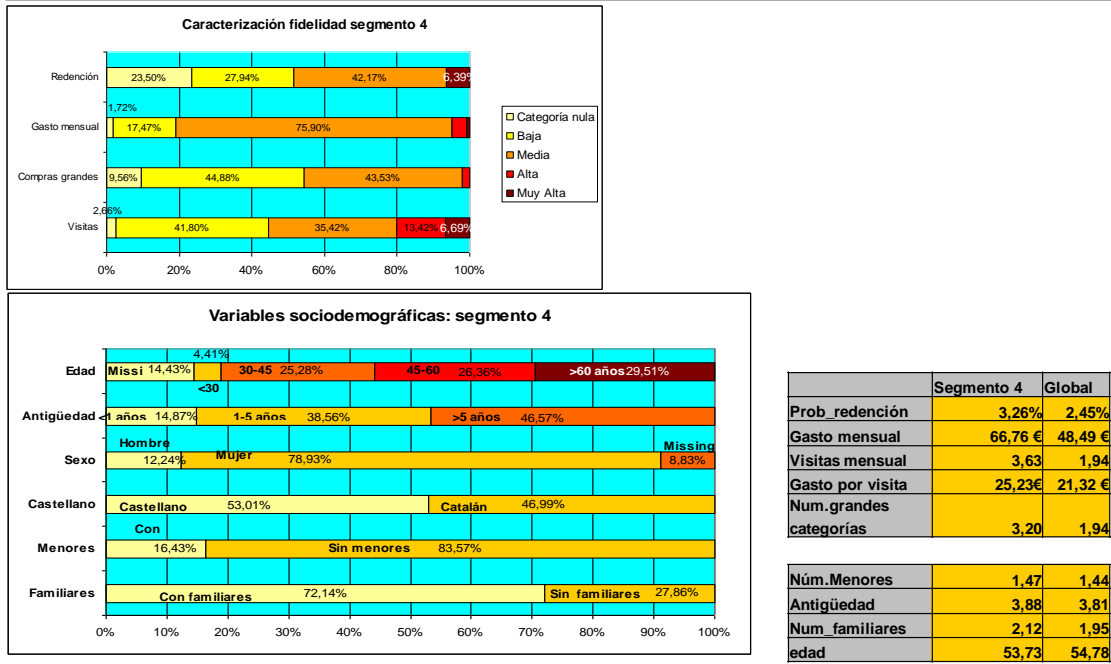
El porcentaje de clientes con menores de edad es superior a la media: es el segundo grupo con mayor porcentaje de edad.

Además el porcentaje de castellanoparlantes sigue siendo superior (57 % de castellano parlantes frente al 43%) y el porcentaje de clientes que viven con familiares sigue siendo elevado.

7.3.2.4 DESCRIPCIÓN SEGMENTO 4 :“Gasto medio con redención media”

Tamaño: 14.71%

Figura 7.3.2.4. Caracterización segmento 4

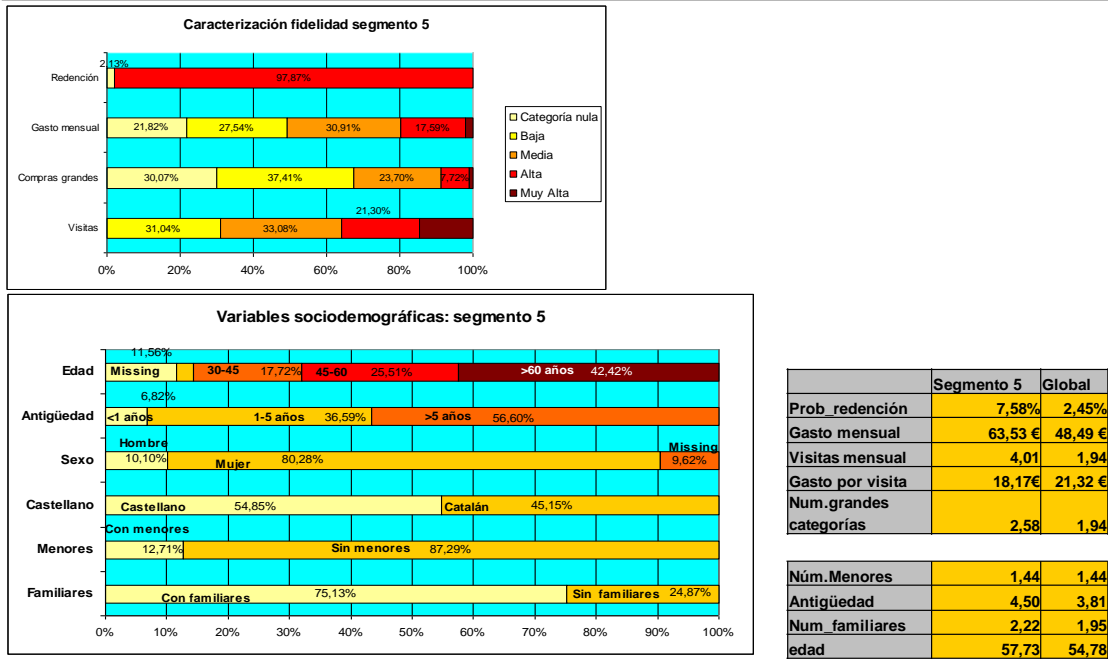


Este grupo se caracteriza por tener una sensibilidad media a los cupones, tienen un gasto medio y solamente compran en promedio en 3 secciones. Tienen una edad similar a la distribución global. Su antigüedad es similar a la media. Existe una ligera mayoría de castellanoparlantes (53% de castellanohablantes).

7.3.2.5 DESCRIPCIÓN SEGMENTO 5 “Los más sensibles al cupón”

Tamaño: 6.57%

Figura 7.3.2.5. Caracterización segmento 5



Grupo caracterizado por un nivel de redención alta (“los amantes de los cupones”) y un gasto medio. Además el gasto por visita es bastante reducido.

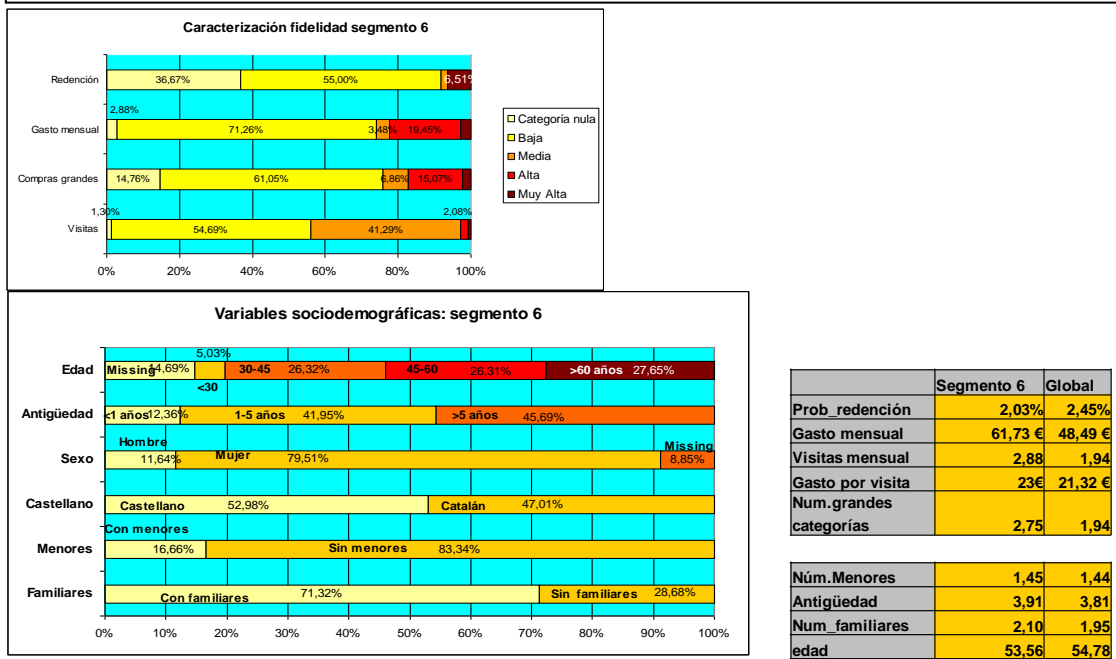
Es el grupo de mayor edad y es el grupo que menos porcentaje de menores de edad tiene.

Además se aprecia que el porcentaje de clientes que viven con familiares es elevado y tienen una antigüedad más elevada que el grupo 4 y 6.

7.3.2.6 DESCRIPCIÓN SEGMENTO 6 “Gasto medio con escasa sensibilidad al cupón”

Tamaño: 12.70%

Figura 7.3.2.6. Caracterización segmento 6



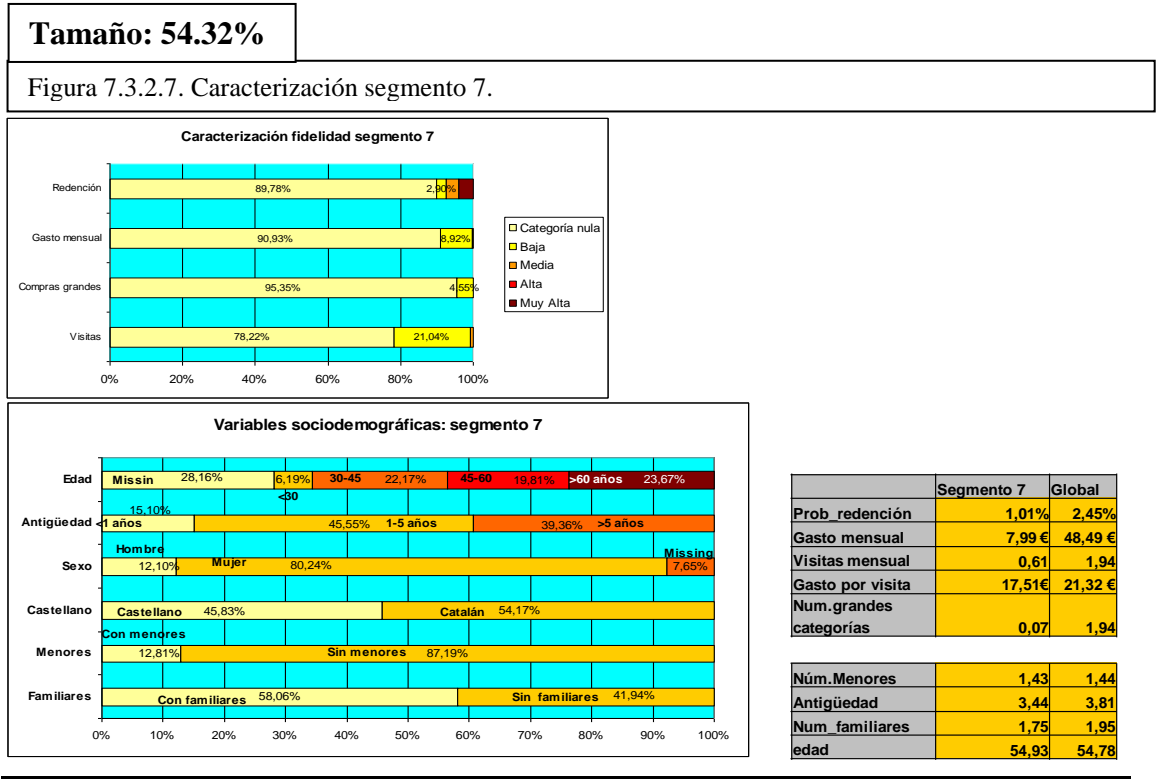
Este segmento se describe fundamentalmente por tener una escasa sensibilidad al cupón y con un gasto medio-bajo.

Realizan menos visitas y son más jóvenes que el segmento 5.

Además tienen un porcentaje de menores de edad superior al segmento 5.

En resumen, es un segmento parecido al segmento 5 pero con escasa redención.

7.3.2.7 DESCRIPCIÓN SEGMENTO 7 “Los menos fieles”



Este segmento es el más numeroso, representa el 54,32% de todos los clientes. Como se puede observar en los gráficos anteriores es el segmento menos fiel ya que gasta menos de 20€ al mes, no hace uso de los cupones, realiza menos de una visita al mes y no realiza ninguna compra grande en ninguna sección.

Este conglomerado destaca por tener un porcentaje de catalanoparlantes superior. Es el único grupo donde la mayoría de los clientes utiliza el catalán como primera lengua (54% de clientes que utilizan el catalán como primera lengua).

Respecto a la edad destaca un porcentaje elevado en el grupo de los menores de 30 años que es superior comparando con otros grupos. Además el porcentaje de valores faltantes es bastante elevado, lo que nos indica que la distribución de missings no es aleatoria sino que se concentra en este segmento.

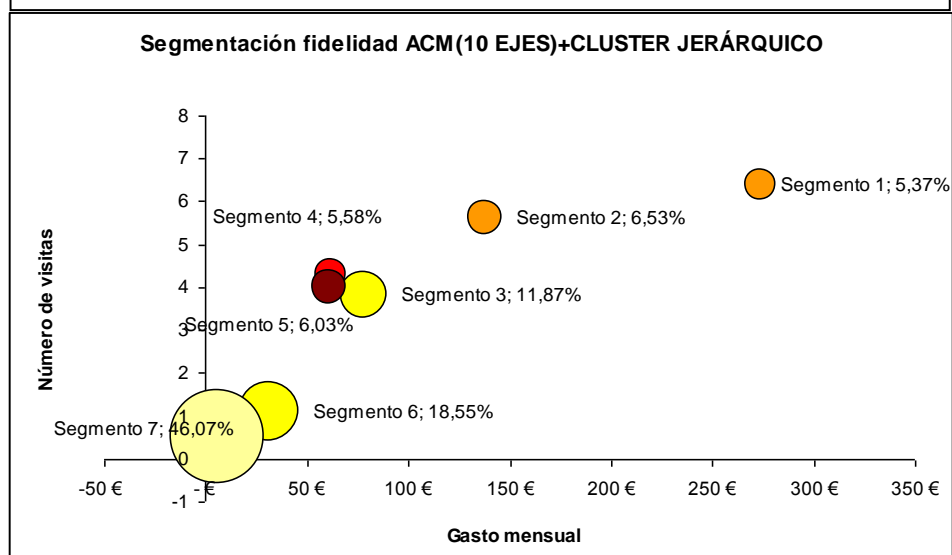
Es digno de mencionar que este es el grupo con menos antigüedad. Además es el único grupo donde el porcentaje de clientes que viven sin familiares es elevado comparado con la media.

7.3.3 RESULTADOS SEGMENTACIÓN CON ALGORITMO JERÁRQUICO

En este apartado se muestran los resultados obtenidos después de aplicar el algoritmo del cluster jerárquico según el criterio de Ward sobre las coordenadas factoriales procedentes de un ACM. Como el algoritmo de cluster jerárquico no es viable computacionalmente con 188.275 individuos se ha seleccionado una muestra que representa el 10% de toda la población mediante un muestreo aleatorio simple. Con esta muestra se ha realizado primero un Análisis de Correspondencias Múltiple y con las 10 coordenadas factoriales obtenidas se aplica el cluster jerárquico.

Esta segunda segmentación se ha realizado con el software de estadística multivariante SPAD.

Figura 7.3.3.1. Representación de los segmentos obtenidos en función de su gasto mensual y número de visitas.



El gráfico 7.3.3.1 nos ayuda a entender de forma general cada segmento según su grado de fidelidad. Así, se aprecia que existe un segmento (segmento número 1) que es el segmento más fiel: los que más gastan, más visitan y los que compran la totalidad de sus compras en el supermercado.

Además existe el segmento 2 con un gasto mensual y un número de visitas elevado, aunque este segmento es un grupo con menos nivel de fidelidad ya que no realiza compras importantes en todas las secciones y por lo tanto su gasto mensual es inferior al segmento 1.

También se aprecia que con un gasto medio existen 3 segmentos diferenciados por su nivel de redención. Así el segmento 4 se caracteriza por tener una muy elevada sensibilidad a los cupones, el segmento 5 tiene una sensibilidad media-alta y el segmento 3 se caracteriza por tener un nivel de sensibilidad reducido al uso de los cupones. Además hay que señalar la existencia del segmento 6 caracterizado por un gasto medio-bajo y el segmento 7 que es el menos fiel y es el que representa al 46% de todos los clientes.

Figura 7.3.3.2. Tabla resumen de cada segmento (Segmentación de fidelidad con cluster jerárquico).

Segmento	Tamaño	Gasto mensual	P(redención)	Visitas	Compras Grandes	Etiqueta
Segmento 1	5,37%	273,86 €	4,07%	6,4	9,8	"Los más fieles"
Segmento 2	6,53%	138,04 €	3,71%	5,6	7,2	"Grado de fidelidad alto, pero selectivos"
Segmento 3	11,87%	78,31 €	2,00%	3,8	4,0	"Gasto medio sin sensibilidad a los cuones"
Segmento 4	5,58%	61,63 €	16,40%	4,3	2,5	"Gasto medio con un alto nivel de redención"
Segmento 5	6,03%	61,42 €	7,72%	4,0	2,4	"Gasto medio con un nivel de redención medio-alto"
Segmento 6	18,55%	31,50 €	1,39%	1,1	1,1	"Gasto medio-bajo con escasa sensibilidad"
Segmento 7	46,07%	5,98 €	0,11%	0,5	0,0	"Los menos fieles"

En el gráfico 7.3.3.3 se proyectan los diferentes segmentos obtenidos en un plano factorial junto con las modalidades activas que forman parte del Análisis de Correspondencias Múltiples.

Las conclusiones son idénticas que las descritas anteriormente: el segmento 7 se asocia con aquellas modalidades relacionadas con un nivel de fidelidad nulo (gasto mensual inferior a 20€, nivel de redención nulo, menos de una visita al mes y no realiza ninguna compra grande), el segmento 1 se asocia con aquellas modalidades referentes a un mayor gasto mensual.

Es interesante ver la posición del segmento 4 y el segmento 5 (los segmentos más sensibles a los cupones). En el gráfico se aprecia que estos dos conglomerados están más próximos a la modalidad más alta de redención (redención mayor del 10%), especialmente el segmento 4.

Además en la figura 7.3.3.4 se presenta un resumen de los segmentos obtenidos y a continuación se presentan la descripción de cada segmento que proporciona SPAD.

Gráfico 7.3.3.3. Proyección de los segmentos obtenidos en un plano factorial.

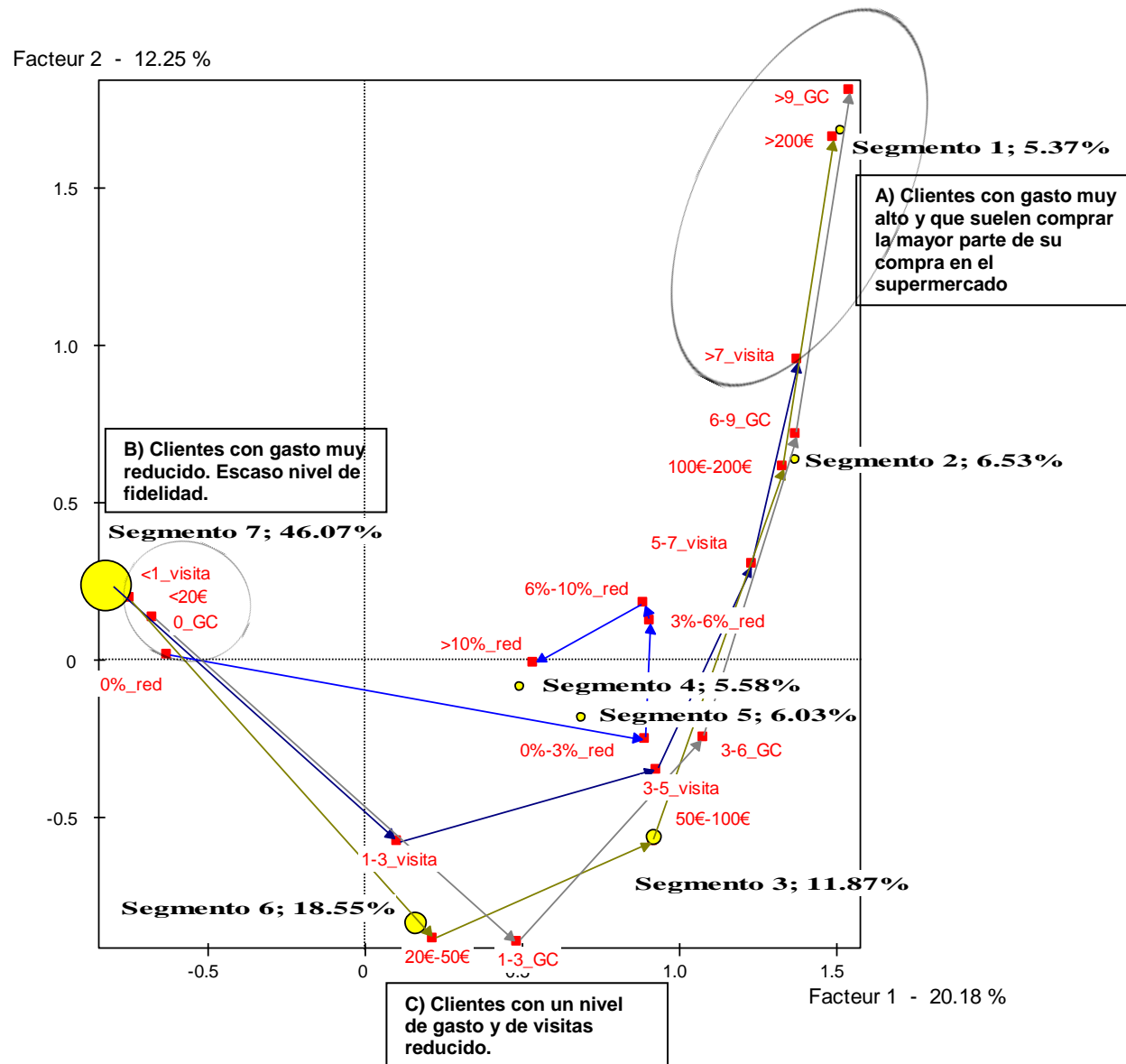


Figura 7.3.3.4. Resumen segmentos obtenidos

GASTO ALTO

SEGMENTO 1 (5.37%)

Es el segmento más fiel: realiza casi todas sus compras en el supermercado.

- Gasto superior a 200€ y número de visitas altas.
- Nivel de redención medio.
- Predominio en el grupo de edad de 45-60 años.
- Antigüedad superior a 5 años. Es el grupo con mayor antigüedad.
- Hablan castellano como primera lengua.
- Viven con menores y con familiares en el hogar.

SEGMENTO 2 (6.53%)

Son los segundos más visitantes y los segundos que más gastan. Son clientes selectivos que no realizan compras considerables en todas las secciones.

- Nivel de redención medio.
- Predominio en el grupo de edad de 45-60 años.
- Antigüedad superior a 5 años.
- Hablan castellano como primera lengua.

GASTO MEDIO

SEGMENTO 3 (11,87%)

-Gasto medio sin sensibilidad al cupón.

- Número de visitas medio.
- Segmento más joven.
- Tiene importancia el grupo con una edad de 30-45 años.
- Realizan menos visitas que el grupo 4 y 5 y son más jóvenes.
- Porcentaje de menores de edad superior a la media.

SEGMENTO 4 (5,58%)

-"Los amantes de los cupones", cuentan con un nivel de redención superior al 10%.

- Es el segmento con mayor edad.
- Predominio del grupo de mayores de 60 años.
- Dentro de los segmentos con un gasto medio (segmento 3, 4 y 5) es el que más visitas realiza.
- Viven con familiares en el hogar.

SEGMENTO 5 (6,03%)

-Gasto medio con una sensibilidad media-alta.

- Es el segundo segmento con mayor edad.
- Predominio del grupo de mayores de 60 años.
- Viven con familiares en el hogar.
- Hablan castellano

GASTO REDUCIDO

SEGMENTO 6 (18.55%)

-Gasto bajo con escasa sensibilidad a los cupones.

- Gasto mensual bajo comprendido entre 20€-50€.
- Realizan compras "considerables" en una o tres secciones
- Visitan el supermercado entre 1 y 3 veces al mes.
- El porcentaje de clientes que viven con familiares es superior a la media.

SEGMENTO 7 (46.07%)

Segmento menos fiel.

- Segmento más numeroso: el 54,32%.
- Predominio de catalanoparlantes.
- Destaca el grupo de menores de 30 años.
- Viven sin familiares.
- Viven sin menores de edad.
- Grupo con menor antigüedad con la tarjeta de fidelidad.

7.3.3.1 DESCRIPCIÓN SEGMENTO 1 “Los más fieles”

Tamaño: 5.37%

Figura 7.3.3.1. Salida de Spad: caracterización segmento 1.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_gasto_mensual	200€-1000€	81,75	4,90	89,59	52,53	0,000	490
cat_grandes_categorias	>9 GC	72,25	4,01	96,76	50,26	0,000	401
cat_visitas	>7 visita	42,46	7,16	31,84	24,02	0,000	716
cat_visitas	5-7 visita	29,80	8,19	19,54	15,15	0,000	819
cat_redencion	3%-6% red	38,18	13,25	15,47	14,99	0,000	1325
cat_redencion	0%-3% red	38,55	15,72	13,17	13,19	0,000	1572
cat_grandes_categorias	6-9 GC	22,72	8,09	15,08	10,78	0,000	809
cat_antigu	antig>5	56,24	35,76	8,45	9,89	0,000	3576
cat_edad	45-60	38,36	23,75	8,67	7,73	0,000	2375
IDIOMA	Castellano	66,11	50,10	7,09	7,65	0,000	5010
cat_redencion	6%-10% red	16,01	7,92	10,86	6,35	0,000	792
cat_gasto_mensual	100€-200€	18,06	10,24	9,47	5,60	0,000	1024
cat_fam	si familiares	74,86	65,14	6,17	4,93	0,000	6514
cat_visitas	3-5 visita	21,42	14,22	8,09	4,60	0,000	1422
cat_menores	missing	13,78	8,37	8,84	4,26	0,000	837
SEXO	sex missing	12,85	8,73	7,90	3,23	0,001	873
cat_menores	Si menores	9,12	5,70	8,60	3,21	0,001	570
cat_edad	60-90	22,53	27,03	4,48	-2,40	0,008	2703
cat_edad	<30	1,68	5,14	1,75	-4,12	0,000	514
cat_fam	no familiares	24,77	34,73	3,83	-5,06	0,000	3473
cat_edad	edad missing	12,10	21,16	3,07	-5,54	0,000	2116
cat_menores	No menores	77,10	85,93	4,82	-5,62	0,000	8593
IDIOMA	Catalan	33,89	49,90	3,65	-7,65	0,000	4990
cat_grandes_categorias	3-6 GC	1,86	10,70	0,93	-8,05	0,000	1070
cat_antigu	antig<1	9,50	28,09	1,82	-10,80	0,000	2809
cat_visitas	1-3 visita	6,33	26,40	1,29	-12,24	0,000	2640
cat_gasto_mensual	50€-100€	0,19	15,04	0,07	-12,81	0,000	1504
cat_grandes_categorias	1-3 GC	1,12	20,23	0,30	-14,02	0,000	2023
cat_gasto_mensual	20€-50€	0,00	18,20	0,00	-14,68	0,000	1820
cat_visitas	<1 visita	0,00	44,03	0,00	-25,28	0,000	4403
cat_redencion	0% red	3,17	57,31	0,30	-28,04	0,000	5731
cat_gasto_mensual	<20€	0,00	51,62	0,00	-28,35	0,000	5162
cat_grandes_categorias	0 GC	2,05	56,97	0,19	-28,73	0,000	5697

Es el segmento más fiel: realiza casi todas sus compras en el supermercado y se caracteriza por tener un gasto superior a 200€ y por un número de visitas altas.

En cuanto a la redención no está tan claro: podemos decir que en general se caracteriza por tener un nivel de redención medio.

Respecto a la edad existe un predominio en el grupo de edad de 45-60 años y además el porcentaje de menores de 30 años es nulo.

En lo que a la antigüedad se refiere se caracteriza por tener una antigüedad superior a 5 años. Es el grupo con mayor antigüedad.

Es el grupo que cuenta con el mayor porcentaje de castellanohablantes. Además se aprecia que este segmento tiene el mayor porcentaje de clientes que viven con menores de hogar y con familiares.

7.3.3.2 DESCRIPCIÓN SEGMENTO 2 “Grado de fidelidad alto pero no compran en todas las secciones”

Tamaño: 6.53%

Figura 7.3.3.2. Salida de Spad: caracterización segmento 2.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_gasto_mensual	100€-200€	94,03	10,24	59,96	54,22	0,000	1024
cat_grandes_categorias	6-9 GC	80,09	8,09	64,65	49,04	0,000	809
cat_visitas	5-7 visita	33,38	8,19	26,62	19,32	0,000	819
cat_redencion	3%-6% red	41,19	13,25	20,30	18,51	0,000	1325
cat_visitas	>7 visita	27,11	7,16	24,72	16,39	0,000	716
cat_redencion	0%-3% red	39,05	15,72	16,22	14,98	0,000	1572
cat_visitas	3-5 visita	27,11	14,22	12,45	8,89	0,000	1422
cat_redencion	6%-10% red	15,77	7,92	13,01	6,87	0,000	792
cat_fam	si familiares	75,80	65,14	7,60	6,03	0,000	6514
cat_edad	45-60	32,31	23,75	8,88	5,11	0,000	2375
cat_antigu	antig>5	44,41	35,76	8,11	4,67	0,000	3576
IDIOMA	Castellano	55,74	50,10	7,27	2,95	0,002	5010
cat_antigu	antig(1-5)	41,19	36,15	7,44	2,71	0,003	3615
cat_menores	Si menores	8,27	5,70	9,47	2,71	0,003	570
cat_grandes_categorias	3-6 GC	13,94	10,70	8,50	2,62	0,004	1070
SEXO	sex missing	11,64	8,73	8,71	2,56	0,005	873
IDIOMA	Catalan	44,26	49,90	5,79	-2,95	0,002	4990
cat_edad	<30	2,60	5,14	3,31	-3,18	0,001	514
cat_fam	no familiares	24,04	34,73	4,52	-6,05	0,000	3473
cat_edad	edad missing	11,49	21,16	3,54	-6,62	0,000	2116
cat_grandes_categorias	>9 GC	0,00	4,01	0,00	-7,05	0,000	401
cat_gasto_mensual	50€-100€	5,97	15,04	2,59	-7,36	0,000	1504
cat_gasto_mensual	200€-1000€	0,00	4,90	0,00	-7,88	0,000	490
cat_antigu	antig<1	14,40	28,09	3,35	-8,51	0,000	2809
cat_redencion	>10% red	0,00	5,80	0,00	-8,64	0,000	580
cat_visitas	1-3 visita	12,10	26,40	2,99	-9,18	0,000	2640
cat_grandes_categorias	1-3 GC	3,37	20,23	1,09	-12,99	0,000	2023
cat_gasto_mensual	20€-50€	0,00	18,20	0,00	-16,28	0,000	1820
cat_visitas	<1 visita	0,31	44,03	0,05	-27,59	0,000	4403
cat_redencion	0% red	3,98	57,31	0,45	-30,46	0,000	5731
cat_grandes_categorias	0 GC	2,60	56,97	0,30	-31,39	0,000	5697
cat_gasto_mensual	<20€	0,00	51,62	0,00	-31,44	0,000	5162

Son los segundos más visitantes y los segundos que más gastan.

Son clientes selectivos que no compran todo en el supermercado.

Tienen un gasto mensual alto. Además se caracterizan por tener un nivel de redención medio.

Respecto a la edad destaca el grupo con una edad comprendida entre 45-60 años.

Existe un predominio de clientes que hablan castellano y que cuentan con una antigüedad superior a 5 años, aunque en menor medida que en el segmento 1.

7.3.3.3 DESCRIPCIÓN SEGMENTO 3 “Gasto medio sin sensibilidad a los cupones

Tamaño: 11.87%

Figura 7.3.3.3. Salida de Spad: caracterización segmento 3.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_gasto_mensual	50€-100€	88,04	15,04	69,48	63,15	0,000	1504
cat_grandes_categorias	3-6_GC	56,02	10,70	62,15	43,43	0,000	1070
cat_redencion	0%-3%_red	43,89	15,72	33,14	25,09	0,000	1572
cat_visitas	3-5_visita	41,03	14,22	34,25	24,65	0,000	1422
cat_redencion	3%-6%_red	32,01	13,25	28,68	18,11	0,000	1325
cat_grandes_categorias	1-3_GC	32,77	20,23	19,23	10,84	0,000	2023
cat_visitas	5-7_visita	14,07	8,19	20,39	7,26	0,000	819
cat_visitas	1-3_visita	35,21	26,40	15,83	7,12	0,000	2640
cat_fam	si_familiares	72,28	65,14	13,17	5,55	0,000	6514
cat_edad	30-45	28,81	22,92	14,92	5,00	0,000	2292
IDIOMA	Castellano	55,27	50,10	13,09	3,76	0,000	5010
cat_edad	45-60	27,89	23,75	13,94	3,48	0,000	2375
cat menores	Si menores	7,83	5,70	16,32	3,18	0,001	570
cat menores	missing	10,28	8,37	14,58	2,42	0,008	837
IDIOMA	Catalan	44,73	49,90	10,64	-3,76	0,000	4990
cat menores	No menores	81,89	85,93	11,31	-4,10	0,000	8593
cat_fam	no familiares	27,46	34,73	9,39	-5,66	0,000	3473
cat_edad	edad missing	13,98	21,16	7,84	-6,68	0,000	2116
cat_gasto_mensual	200€-1000€	1,18	4,90	2,86	-7,27	0,000	490
cat_grandes_categorias	>9_GC	0,00	4,01	0,00	-9,86	0,000	401
cat_redencion	>10%_red	0,00	5,80	0,00	-12,03	0,000	580
cat_redencion	6%-10%_red	0,00	7,92	0,00	-14,22	0,000	792
cat_gasto_mensual	20€-50€	1,18	18,20	0,77	-19,84	0,000	1820
cat_redencion	0%_red	24,09	57,31	4,99	-24,75	0,000	5731
cat_visitas	<1_visita	2,53	44,03	0,68	-34,92	0,000	4403
cat_grandes_categorias	0_GC	4,72	56,97	0,98	-41,19	0,000	5697
cat_gasto_mensual	<20€	0,00	51,62	0,00	-43,43	0,000	5162

Es un segmento caracterizado por un gasto medio pero sin sensibilidad al cupón, es decir, con un nivel de redención reducido.

Tiene un promedio de visitas medio y es un segmento más joven, adquiriendo importancia el grupo con una edad de 30-45 años.

Realizan menos visitas que el segmento 4 y 5 y son más jóvenes que éstos.

Además tienen un porcentaje de menores de edad superior a la media.

7.3.3.4 DESCRIPCIÓN SEGMENTO 4 “Gasto medio con mucha sensibilidad a los cupones”

Tamaño: 5.58%

Figura 7.3.3.4. Salida de Spad: caracterización segmento 4.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_redencion	>10%_red	100,00	5,80	96,21	64,05	0,000	580
cat_edad	60-90	44,27	27,03	9,14	8,96	0,000	2703
cat_visitas	>7_visita	17,74	7,16	13,83	8,57	0,000	716
cat_visitas	5-7_visita	17,92	8,19	12,21	7,58	0,000	819
cat_gasto_mensual	100€-200€	18,28	10,24	9,96	5,87	0,000	1024
cat_grandes_categorias	1-3_GC	30,47	20,23	8,40	5,86	0,000	2023
cat_antigu	antig>5	47,13	35,76	7,35	5,63	0,000	3576
cat_fam	si familiares	75,63	65,14	6,48	5,45	0,000	6514
cat_gasto_mensual	50€-100€	22,94	15,04	8,51	5,03	0,000	1504
cat_grandes_categorias	3-6_GC	16,67	10,70	8,69	4,35	0,000	1070
cat_visitas	3-5_visita	20,43	14,22	8,02	4,07	0,000	1422
cat_grandes_categorias	6-9_GC	12,01	8,09	8,28	3,24	0,001	809
cat_edad	45-60	29,21	23,75	6,86	3,01	0,001	2375
cat menores	missing	5,38	8,37	3,58	-2,68	0,004	837
cat_grandes_categorias	>9_GC	1,61	4,01	2,24	-3,16	0,001	401
cat_antigu	antig(1-5)	28,85	36,15	4,45	-3,70	0,000	3615
cat_fam	no familiares	24,37	34,73	3,92	-5,38	0,000	3473
cat_edad	30-45	12,37	22,92	3,01	-6,44	0,000	2292
cat_edad	edad_missing	10,57	21,16	2,79	-6,71	0,000	2116
cat_gasto_mensual	<20€	34,59	51,62	3,74	-8,29	0,000	5162
cat_grandes_categorias	0_GC	39,25	56,97	3,84	-8,61	0,000	5697
cat_redencion	6%-10%_red	0,00	7,92	0,00	-9,42	0,000	792
cat_redencion	3%-6%_red	0,00	13,25	0,00	-12,52	0,000	1325
cat_visitas	<1_visita	19,00	44,03	2,41	-12,77	0,000	4403
cat_redencion	0%-3%_red	0,00	15,72	0,00	-13,78	0,000	1572
cat_redencion	0%_red	0,00	57,31	0,00	-31,40	0,000	5731

Es el segmento más sensible a los cupones, cuentan con un nivel de redención superior al 10%. Se puede decir que son los **amantes de los cupones**.

Respecto a la edad se puede decir que es el segmento con mayor edad en el que el grupo de 60 a 90 años tiene una gran importancia.

Dentro de los segmentos con un gasto medio (segmento 3, 4 y 5) es el que más visitas realiza.

Está compuesto por un porcentaje elevado de clientes que viven con familiares.

Además el porcentaje de clientes con una antigüedad mayor de 5 años es superior a la media.

7.3.3.5 DESCRIPCIÓN SEGMENTO 5“Gasto medio con un nivel de redención medio alto”

Tamaño: 6.03%

Figura 7.3.3.5. Salida de Spad: caracterización segmento 5.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_redencion	6%-10% red	100,00	7,92	76,14	60,60	0,000	792
cat_grandes_categorias	3-6 GC	30,02	10,70	16,92	13,50	0,000	1070
cat_gasto_mensual	50€-100€	36,15	15,04	14,49	13,24	0,000	1504
cat_grandes_categorias	1-3 GC	34,83	20,23	10,38	8,59	0,000	2023
cat_visitas	3-5 visita	26,87	14,22	11,39	8,36	0,000	1422
cat_visitas	>7 visita	16,42	7,16	13,83	7,92	0,000	716
cat_edad	60-90	40,63	27,03	9,06	7,42	0,000	2703
cat_visitas	5-7 visita	16,58	8,19	12,21	6,91	0,000	819
cat_antigu	antig>5	46,93	35,76	7,91	5,77	0,000	3576
cat_gasto_mensual	100€-200€	14,59	10,24	8,59	3,41	0,000	1024
cat_fam	si familiares	70,98	65,14	6,57	3,10	0,001	6514
cat_gasto_mensual	20€-50€	22,72	18,20	7,53	2,85	0,002	1820
IDIOMA	Castellano	55,39	50,10	6,67	2,64	0,004	5010
IDIOMA	Catalan	44,61	49,90	5,39	-2,64	0,004	4990
cat_fam	no familiares	28,69	34,73	4,98	-3,21	0,001	3473
cat_gasto_mensual	200€-1000€	2,16	4,90	2,65	-3,43	0,000	490
cat_grandes_categorias	6-9 GC	3,15	8,09	2,35	-5,03	0,000	809
cat_grandes_categorias	>9 GC	0,66	4,01	1,00	-5,06	0,000	401
cat_edad	edad missing	12,94	21,16	3,69	-5,32	0,000	2116
cat_antigu	antig<1	17,91	28,09	3,84	-5,93	0,000	2809
cat_redencion	>10% red	0,00	5,80	0,00	-8,27	0,000	580
cat_grandes_categorias	0 GC	31,34	56,97	3,32	-13,05	0,000	5697
cat_redencion	3%-6% red	0,00	13,25	0,00	-13,05	0,000	1325
cat_gasto_mensual	<20€	24,38	51,62	2,85	-14,04	0,000	5162
cat_redencion	0%-3% red	0,00	15,72	0,00	-14,36	0,000	1572
cat_visitas	<1 visita	14,26	44,03	1,95	-16,10	0,000	4403
cat_redencion	0% red	0,00	57,31	0,00	-32,71	0,000	5731

Es un grupo caracterizado por un **gasto medio y que tiene una sensibilidad media-alta al uso de los cupones.**

Es el segundo segmento con mayor edad y para su nivel de gasto tiene un número de visitas considerable.

Además tiene un porcentaje elevado de clientes que viven con familiares .También hay una ligera mayoría de clientes que hablan castellano.

El porcentaje de clientes que cuentan con una antigüedad mayor de 5 años es superior a la media.

En resumen es un grupo muy parecido al segmento 4 pero con un nivel de redención inferior.

7.3.3.6 DESCRIPCIÓN SEGMENTO 6 “Gasto medio-bajo con escasa sensibilidad”

Tamaño: 18.55%

Figura 7.3.3.6. Salida de Spad: caracterización segmento 6.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_gasto_mensual	20€-50€	77,90	18,20	79,40	66,95	0,000	1820
cat_grandes_categorias	1-3_GC	61,99	20,23	56,85	45,35	0,000	2023
cat_visitas	1-3_visita	69,97	26,40	49,17	44,55	0,000	2640
cat_redencion	0%-3%_red	29,06	15,72	34,29	16,37	0,000	1572
cat_redencion	3%-6%_red	20,54	13,25	28,75	9,75	0,000	1325
cat_fam	si_familiares	70,13	65,14	19,97	5,02	0,000	6514
cat_visitas	3-5_visita	17,84	14,22	23,28	4,80	0,000	1422
cat_edad	30-45	25,44	22,92	20,59	2,81	0,002	2292
cat_edad	60-90	29,27	27,03	20,09	2,37	0,009	2703
cat_fam	no_familiares	29,81	34,73	15,92	-4,95	0,000	3473
cat_redencion	0%_red	50,40	57,31	16,31	-6,61	0,000	5731
cat_edad	edad_missing	15,09	21,16	13,23	-7,28	0,000	2116
cat_visitas	5-7_visita	3,83	8,19	8,67	-8,18	0,000	819
cat_visitas	>7_visita	1,40	7,16	3,63	-12,38	0,000	716
cat_grandes_categorias	>9_GC	0,00	4,01	0,00	-12,71	0,000	401
cat_gasto_mensual	200€-1000€	0,16	4,90	0,61	-13,22	0,000	490
cat_redencion	>10%_red	0,00	5,80	0,00	-15,45	0,000	580
cat_grandes_categorias	3-6_GC	1,62	10,70	2,80	-16,50	0,000	1070
cat_gasto_mensual	50€-100€	3,83	15,04	4,72	-16,91	0,000	1504
cat_redencion	6%-10%_red	0,00	7,92	0,00	-18,24	0,000	792
cat_grandes_categorias	6-9_GC	0,00	8,09	0,00	-18,45	0,000	809
cat_gasto_mensual	100€-200€	0,49	10,24	0,88	-19,07	0,000	1024
cat_grandes_categorias	0_GC	36,39	56,97	11,85	-19,75	0,000	5697
cat_gasto_mensual	<20€	17,63	51,62	6,33	-33,44	0,000	5162
cat_visitas	<1_visita	6,95	44,03	2,93	-38,85	0,000	4403

Este segmento se identifica por las siguientes características:

- Gasto mensual reducido comprendido entre 20€-50€.
- Compran solo en una o tres secciones y visitan el supermercado entre 1 y 3 veces al mes.
- El porcentaje de clientes que viven con familiares es superior a la media.

7.3.3.7 DESCRIPCIÓN SEGMENTO 7 “Los menos fieles”

Tamaño: 46.07%

Figura 7.3.3.7. Salida de Spad: caracterización segmento 7.

Variable label	Characteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
cat_gasto_mensual	<20€	97,57	51,62	87,08	93,58	0,000	5162
cat_visitas	<1 visita	87,91	44,03	91,98	87,81	0,000	4403
cat_grandes_categorias	0 GC	98,33	56,97	79,52	85,12	0,000	5697
cat_redencion	0% red	96,96	57,31	77,94	80,73	0,000	5731
cat_edad	edad missing	30,24	21,16	65,83	20,58	0,000	2116
cat_fam	no familiares	43,30	34,73	57,44	16,63	0,000	3473
cat_antigu	antig<1	34,58	28,09	56,71	13,31	0,000	2809
IDIOMA	Catalan	55,09	49,90	50,86	9,58	0,000	4990
cat_menores	No menores	88,67	85,93	47,54	7,30	0,000	8593
cat_edad	<30	6,25	5,14	56,03	4,60	0,000	514
SEXO	sex missing	7,81	8,73	41,24	-2,97	0,001	873
cat_edad	30-45	21,42	22,92	43,06	-3,27	0,001	2292
cat_menores	missing	7,14	8,37	39,31	-4,08	0,000	837
cat_menores	Si menores	4,19	5,70	33,86	-6,04	0,000	570
cat_edad	60-90	23,10	27,03	39,36	-8,19	0,000	2703
IDIOMA	Castellano	44,91	50,10	41,30	-9,58	0,000	5010
cat_edad	45-60	18,99	23,75	36,84	-10,37	0,000	2375
cat_antigu	antig>5	29,17	35,76	37,58	-12,73	0,000	3576
cat_fam	si familiares	56,61	65,14	40,04	-16,53	0,000	6514
cat_grandes_categorias	>9 GC	0,00	4,01	0,00	-22,40	0,000	401
cat_gasto_mensual	200€-1000€	0,04	4,90	0,41	-24,40	0,000	490
cat_redencion	>10% red	0,00	5,80	0,00	-27,17	0,000	580
cat_visitas	>7 visita	0,04	7,16	0,28	-29,96	0,000	716
cat_redencion	6%-10% red	0,00	7,92	0,00	-32,05	0,000	792
cat_visitas	5-7 visita	0,07	8,19	0,37	-32,07	0,000	819
cat_grandes_categorias	6-9 GC	0,02	8,09	0,12	-32,21	0,000	809
cat_visitas	1-3 visita	11,20	26,40	19,55	-32,86	0,000	2640
cat_redencion	3%-6% red	1,95	13,25	6,79	-33,60	0,000	1325
cat_gasto_mensual	100€-200€	0,00	10,24	0,00	-36,81	0,000	1024
cat_grandes_categorias	3-6 GC	0,00	10,70	0,00	-37,70	0,000	1070
cat_visitas	3-5 visita	0,78	14,22	2,53	-40,09	0,000	1422
cat_gasto_mensual	20€-50€	2,34	18,20	5,93	-41,48	0,000	1820
cat_redencion	0%-3% red	1,09	15,72	3,18	-41,55	0,000	1572
cat_gasto_mensual	50€-100€	0,04	15,04	0,13	-45,23	0,000	1504
cat_grandes_categorias	1-3 GC	1,65	20,23	3,76	-47,32	0,000	2023

El segmento 7 es el menos fiel ya que gasta menos de 20€ al mes, no hace uso de los cupones, realiza menos de una visita al mes y no realiza ninguna compra grande en ninguna sección.

Es el único grupo donde la mayoría de los clientes utiliza el catalán como primera lengua

Respecto a la edad destaca un porcentaje elevado en el grupo de los menores de 30 años que es superior comparado con otros grupos.

Es el grupo con menos antigüedad y es el único grupo donde el porcentaje de clientes que viven sin familiares es elevado comparado con la media.

8.SEGMENTACIÓN DEL “SHARE OF WALLET”

8.1DESCRIPCIÓN DE LA SEGMENTACIÓN

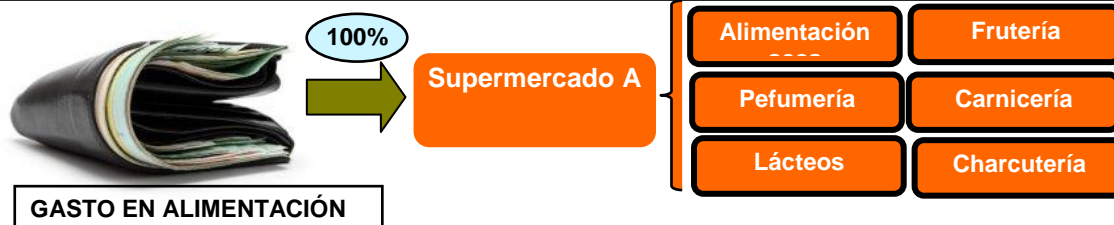
El “Share of Wallet” (SOW) es un concepto de Marketing que hace referencia al porcentaje del gasto total de un cliente que destina a comprar productos en una determinada compañía, es decir, una segmentación del “Share of Wallet” consiste en dividir a los clientes en función de cómo distribuyen su gasto entre los diferentes establecimientos intentando inferir qué clientes realizan la mayor parte de su gasto en un establecimiento y quiénes no.

Figura 8.1.1 Imagen del Share of Wallet



Para un supermercado lo ideal sería que todos sus clientes realizaran todas sus compras en el mismo establecimiento y no realicen compras en otros supermercados competidores.

Figura 8.1.2 Imagen de un cliente que realiza toda su compra en el mismo establecimiento.

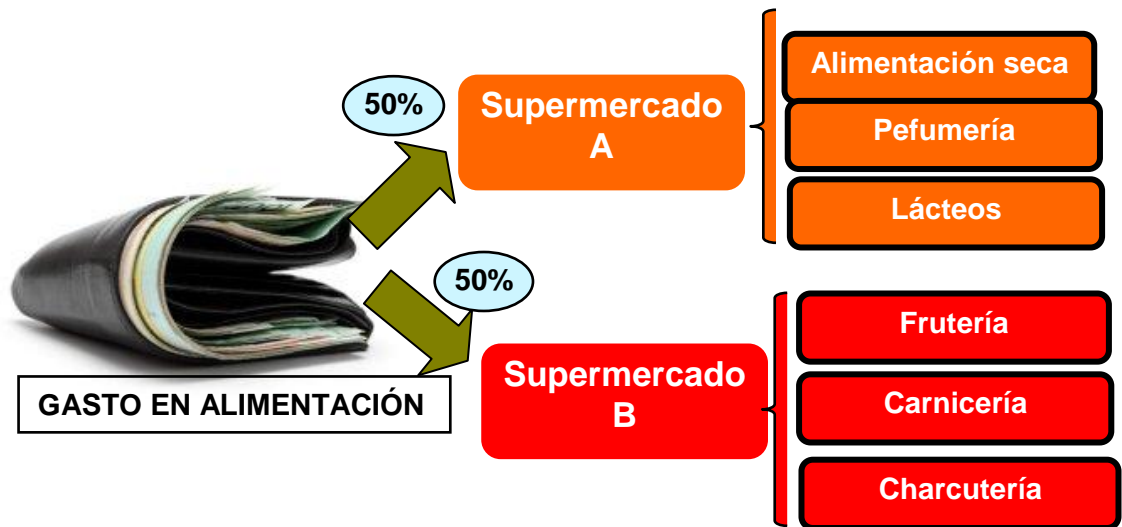


Lo que se pretende con esta segmentación es segmentar a los clientes en función de cómo reparte cada cliente su gasto mensual entre las diferentes secciones en las que se divide el supermercado intentando detectar de esta forma qué conjunto de productos no son adquiridos en el supermercado.

El objetivo planteado es **intentar inferir y suponer** la existencia de tres tipologías de clientes:

- 1) Clientes que tienden a comprar la totalidad de su compra en el mismo establecimiento.
- 2) Clientes que escogen diferentes supermercados para realizar sus compras. A estos clientes que no compran todo en el supermercado sino que reparten su compra en varios establecimientos les denominamos **clientes selectivos**. Para estos clientes selectivos se pretende investigar qué compran o en qué sección destinan la mayor parte de su gasto y además averiguar en qué secciones no compra o realizan compras ocasionales. La siguiente figura ayuda a entender el concepto de un cliente selectivo.

Figura 8.1.3 Esquema de un cliente selectivo que reparte su compra en dos supermercados



- 3) Clientes que realizan compras ocasionales o puntuales en cada sección, es decir, existen clientes que gastan en una sección pero no llegan a realizar una compra importante en dicha sección. Estos clientes que realizan compras que no llegan al umbral establecido en el apartado 2.5 son denominados **clientes ocasionales**.

Además se caracterizará a los diferentes grupos obtenidos con otras variables sociodemográficas y de fidelidad para interpretar y comprender mejor cada segmento.

Este tipo de segmentación se puede englobar como una segmentación de fidelidad. En el apartado anterior definíamos fidelidad con las variables gasto mensual, visitas mensuales, número de compras grandes y redención. Pero la fidelidad puede ser entendida desde el punto de vista del “Share of Wallet”: aquellos clientes que tienden a comprar la totalidad de su compra en el mismo establecimiento se consideran como clientes con un alto nivel de fidelidad. En cambio aquellos clientes que son más selectivos tienen un nivel de fidelidad más reducido ya que no realizan todas sus compras en el supermercado.

La utilidad de este análisis resulta clave para la alta dirección del supermercado o el departamento de Marketing ya que podrá aplicar políticas comerciales personalizadas a cada cliente con el objetivo de fomentar el gasto en aquellas secciones en las que un cliente no gasta.

DATOS DE PARTIDA

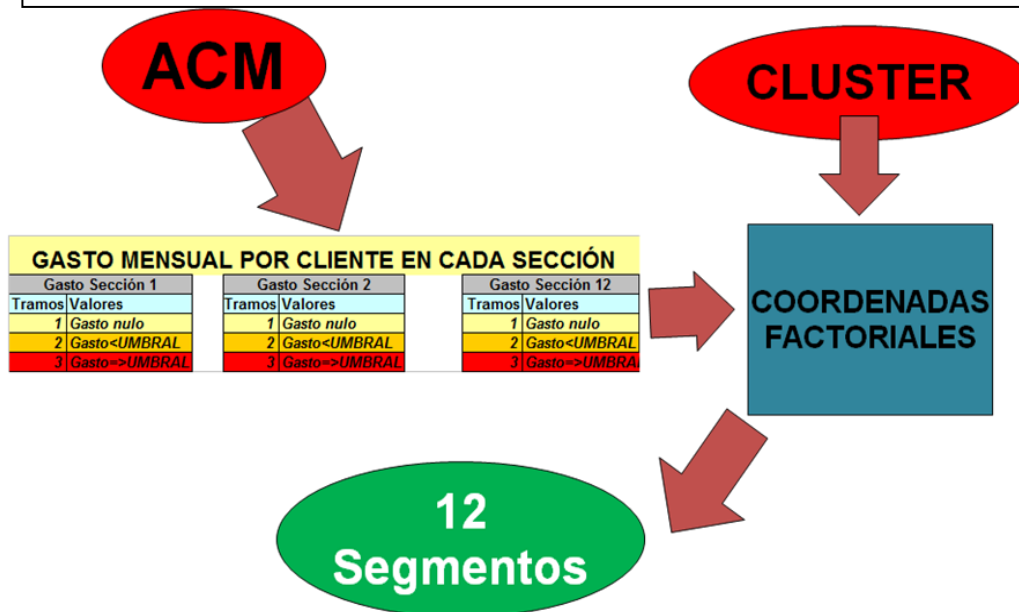
Se parte de un dataset de compras efectuadas por clientes desde enero de 2009 hasta el 30 de septiembre de 2010.

8.2 METODOLOGÍA

La metodología aplicada se puede resumir en las siguientes etapas:

- 1) Desglose del gasto mensual de cada cliente entre las diferentes secciones o categorías de producto que forman el supermercado.
- 2) Categorización de la variable gasto mensual por cada sección en tres modalidades tal como se explica en el apartado 2.6.
- 3) Realización de un Análisis de Correspondencias Múltiples considerando como variables activas las doce variables procedentes del desglose del gasto mensual entre las diferentes secciones y considerando como variables suplementarias las variables de fidelidad de la primera segmentación y las variables sociodemográficas.
- 4) Aplicación del algoritmo cluster K-Means sobre las coordenadas obtenidas de la aplicación del A.C.M y no sobre las variables originales.

Figura 8.2.1 Esquema de la metodología aplicada para la segmentación “Share of Wallet”



En este caso el algoritmo de cluster utilizado ha sido el algoritmo del K-Means. El software empleado ha sido SAS ENTERPRISE GUIDE.

8.3 RESULTADOS

8.3.1 RESULTADOS ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

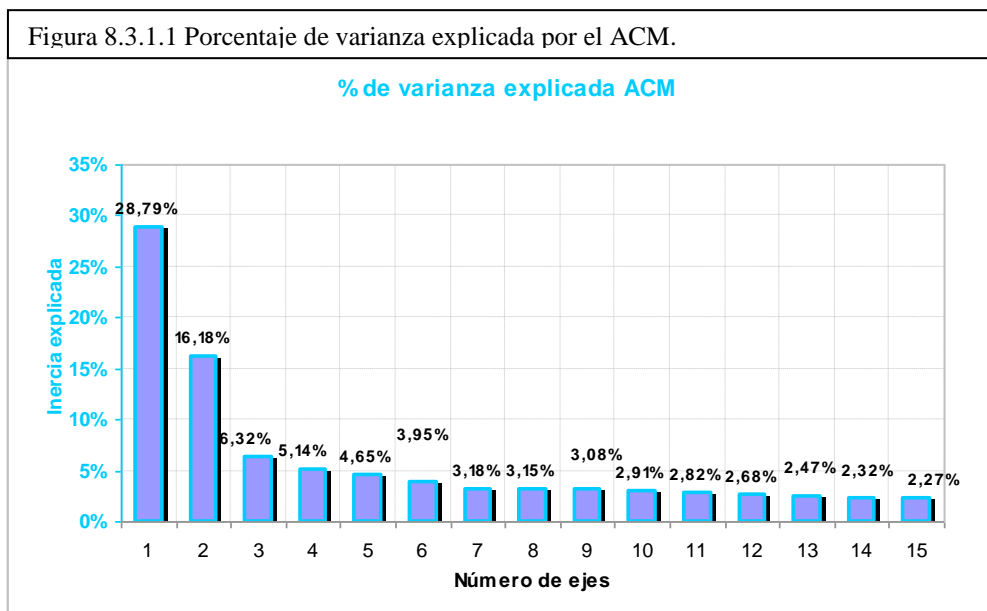
El Análisis de Correspondencias Múltiple que se presenta en este apartado ha sido realizado tanto en SAS como en SPAD. En este apartado se presentará el ACM realizado en SPAD.

Para interpretar el Análisis de Correspondencias Múltiple se sigue el siguiente esquema:

- 1) Análisis de la descomposición de la inercia explicada por cada eje factorial.
- 2) Interpretación gráfica de los ejes factoriales:
 - a. Proyección de las variables activas sobre los ejes factoriales.
 - b. Proyección de las variables suplementarias sobre los ejes factoriales.
 - c. Proyección de la nube de puntos de los individuos junto con las variables activas

1) TABLA DE DESCOMPOSICIÓN DE LA VARIANZA.

Con el siguiente gráfico representamos la descomposición de la varianza explicada por cada eje. En él se puede observar que el primer eje factorial recoge gran parte de la inercia siendo con gran diferencia el eje más diferenciado.



En el histograma se aprecia que los dos primeros ejes recogen aproximadamente el 45 % de toda la inercia, sobresaliendo entre ellos el primer eje con una explicabilidad del 28,79%.

2) INTERPRETACIÓN GRÁFICA DE LOS EJES FACTORIALES.

En el gráfico que se expone a continuación Figura 8.3.1.2 se proyectan las 12 variables de gasto mensual por cliente pertenecientes a cada sección tramificadas en tres modalidades: gasto nulo, compras ocasionales y compras grandes.

Se aprecia claramente en la Figura 8.3.1.2 que en la elipse A situada en la parte inferior izquierda se sitúan las modalidades que hacen referencia a un gasto nulo en cada sección y en la elipse de la derecha aparecen aquellas modalidades que nos indican que el cliente compra todo en esa sección.

El eje1 contrapone clientes que realizan un gasto nulo en cada sección contra aquellos que realizan compras importantes en cada sección, es decir, enfrenta a los que no compran nada en el supermercado contra los que realizan todas sus compras en el mismo establecimiento. De esta forma, el eje 1 se puede interpretar como un eje que hace referencia a la **fidelidad** y se considera un **factor tamaño** ya que opone los individuos que tienen valores altos en todas las variables con los que tienen valores bajos.

La interpretación del segundo eje no está tan clara como en el primer eje debido al efecto Gutman. Se aprecia claramente que tanto en la parte inferior izquierda y en la derecha aparecen las modalidades más extremas y en la parte superior central aparecen las modalidades intermedias.

Con esto se puede decir que la estructura de los datos es casi unidimensional con lo que la interpretación del segundo eje es complicada y en este caso no tiene mucho sentido su interpretación.

Además en la Figura 8.3.1.2 se observa que existen dos secciones (sección carnicería y pescadería) que se escapan de la tendencia general de las demás secciones. La sección carnicería y sobre todo pescadería son las menos competitivas ya que por ejemplo se observa que la modalidad de gasto nulo de la sección pescadería se asocia con aquellas modalidades de compras ocasionales de las demás secciones. Esto quiere decir que los clientes que se sitúan en la zona de la elipse B realizan compras ocasionales en casi todas las secciones excepto en la sección pescadería. También cabe destacar que en la elipse A del gasto nulo en cada sección la sección alimentación seca tiene gran influencia sobre el eje 1 y sobre el eje 2 ya que la modalidad de gasto nulo de alimentación seca es la más extrema. Esto quiere decir que si un individuo no compra en la sección alimentación seca tiende a no comprar en el resto de secciones. Lógicamente como partimos de un conjunto de datos en el que todos los individuos han realizado alguna compra no habrá ningún individuo situado justo en la modalidad de gasto nulo de la sección alimentación seca.

En la figura 8.3.1.3, además de proyectar las doce variables activas (gasto mensual por cliente en cada sección) se hace lo mismo con las variables suplementarias de fidelidad: gasto mensual, visitas mensuales, número de compras grandes, y porcentaje de redención. A simple vista, se aprecia que el gasto mensual, visitas mensuales y el número de compras grandes siguen la tendencia de un efecto Gutman pero la variable redención se escapa de este efecto tal como sucedía en la segmentación de fidelidad mostrada en el apartado anterior. Además merece mención especial la trayectoria de la

variable número de visitas (color rosa) ya que se aprecia que la modalidad más alta de la variable visitas (más de 7 visitas al mes) está más relacionada con la modalidad de gasto entre 100 € y 200€ que con la modalidad más alta de gasto (mayor de 200€). Esto confirma los resultados de la segmentación de fidelidad en el que el grupo que más gasto realizaba se caracterizaba por tener un nivel de gasto por visita muy alto, es decir, tenían un gasto muy elevado comparado con su nivel de visitas.

En la 8.3.1.4, se muestra en detalle ampliando el zoom la proyección de las siguientes variables sociodemográficas: antigüedad, sexo, idioma, si vive con familiares y si vive con menores de edad

Respecto a la variable antigüedad se observa que la modalidad antigüedad inferior a un año está próxima a modalidades de gasto nulo en cada sección, mientras que antigüedades superiores nos indican un mayor grado de fidelidad.

Se obtienen las mismas conclusiones que en el ACM del apartado anterior: aquellos clientes que viven en el hogar con familiares y menores de edad tienen un grado de fidelidad mayor que aquellos que no lo hacen.

Respecto al sexo se aprecia que no hay grandes diferencias en cuanto al comportamiento de los hombres y de las mujeres. Las mujeres están más cerca del baricentro ya que es una modalidad no balanceada. También parece que ligeramente la modalidad catalán se relaciona más con aquellas modalidades que nos indican un menor grado de fidelidad pero no hay grandes diferencias.

En la 8.3.1.5 se presenta la proyección de la variable edad que no ha podido incluirse en el gráfico anterior. En este gráfico la categorización de la variable edad se hace con más detalle en 10 modalidades. Se aprecia claramente que las categorías de menor edad son aquellas que cuentan con un grado de fidelidad más bajo. La tendencia es que el grado de fidelidad aumenta con el tiempo alcanzando su máximo entre el grupo y edad comprendido entre los 45 y 55 años. A partir de esa edad la fidelidad va disminuyendo con el tiempo.

Como complemento se presenta en la figura 8.3.1.6 la proyección de los individuos junto con las variables activas. Del 8.3.1.6 se puede decir que la mayoría de los clientes de la nube de puntos sigue la tendencia del efecto Gutman pero hay una parte de los clientes que se escapa de ese efecto. Los clientes situados en la elipse inferior derecha (elipse C) son aquellos que suelen comprar la mayor parte de sus compras en el supermercado. Así por ejemplo, el cliente 10000008030 se caracteriza por realizar todas las compras en el supermercado, exceptuando para la sección pescadería y frutería donde realiza compras ocasionales. Además en la parte inferior central en la elipse D se aprecia un grupo de clientes que se escapan del efecto Gutman y se caracterizan por realizar compras grandes en algunas secciones y no comprar en otras secciones.

Este conjunto de clientes son los denominados selectivos ya que no realizan toda la compra en el supermercado. Como ejemplo de esta tipología de clientes destaca el cliente con código 1000508229 que realiza compras grandes en Alimentación seca, conservas, cremería, líquidos y panadería pero tienen un gasto nulo en la sección perfumería, frutería, charcutería y pescadería.

Figura 8.3.1.2 Proyección de las variables activas.

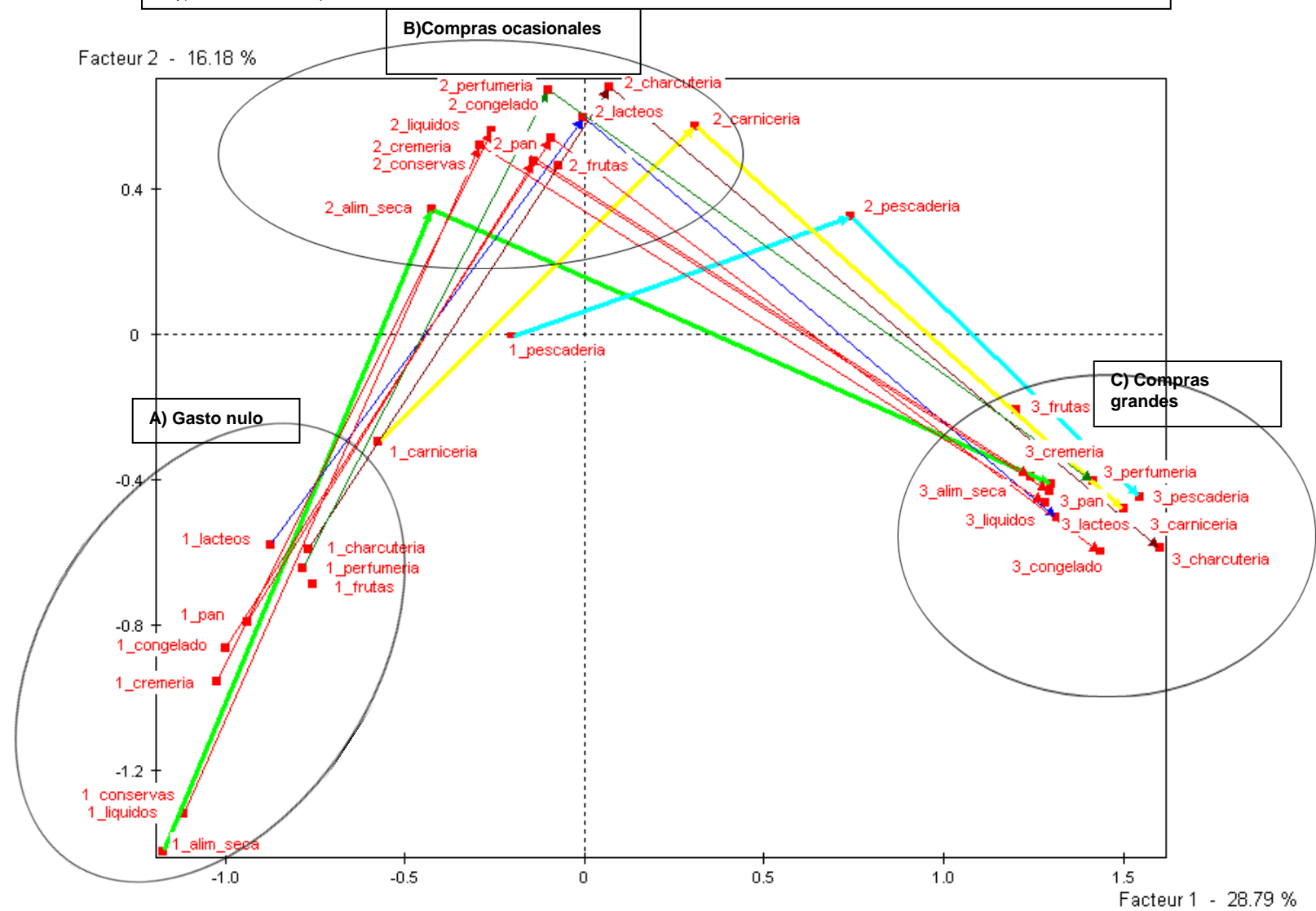


Figura 8.3.1.3: Proyección de las variables suplementarias de fidelidad (gasto, visitas, número de compras grandes y redención)

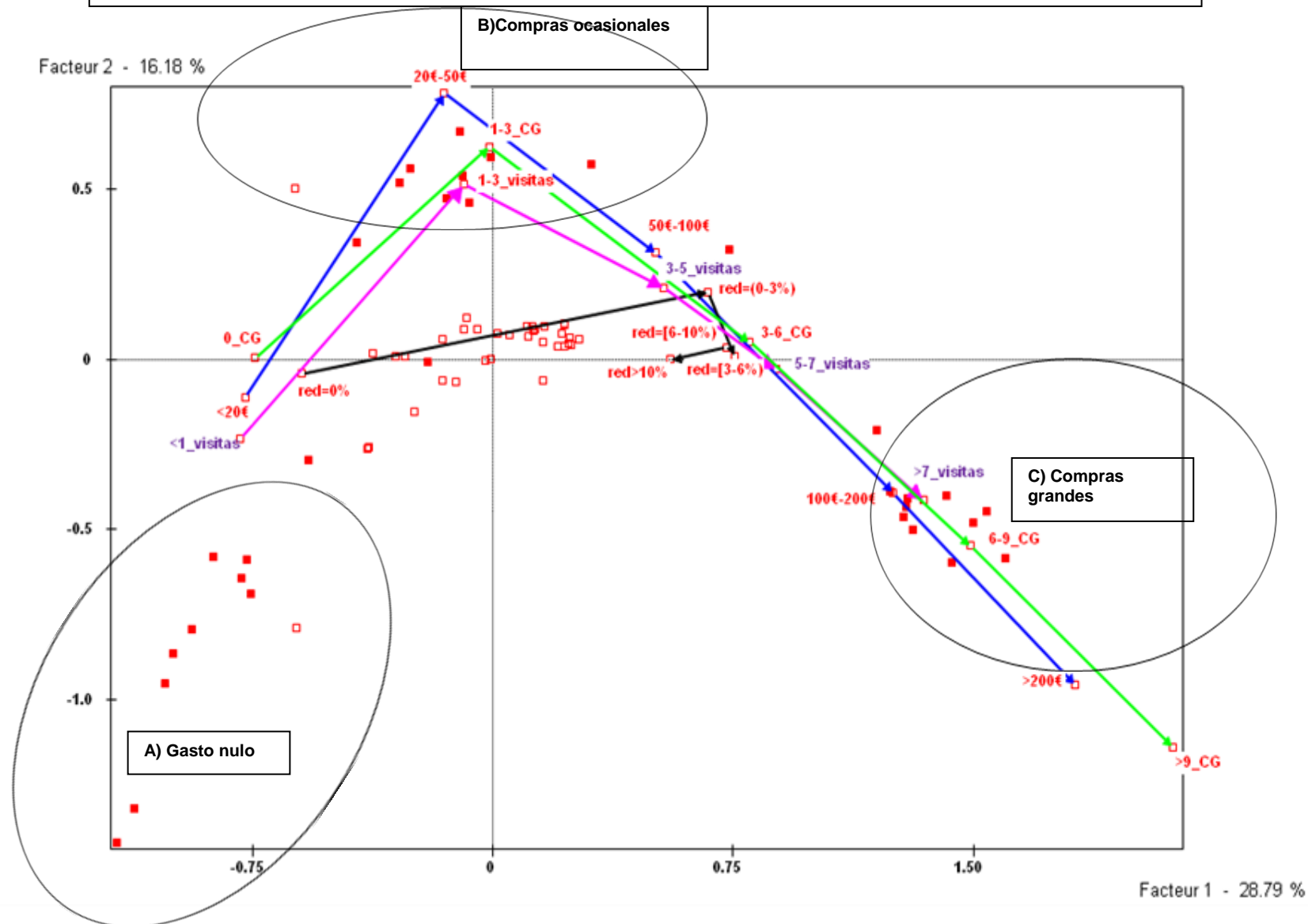


Figura 8.3.1.4: Proyección de las variables suplementarias sociodemográficas excepto la edad.

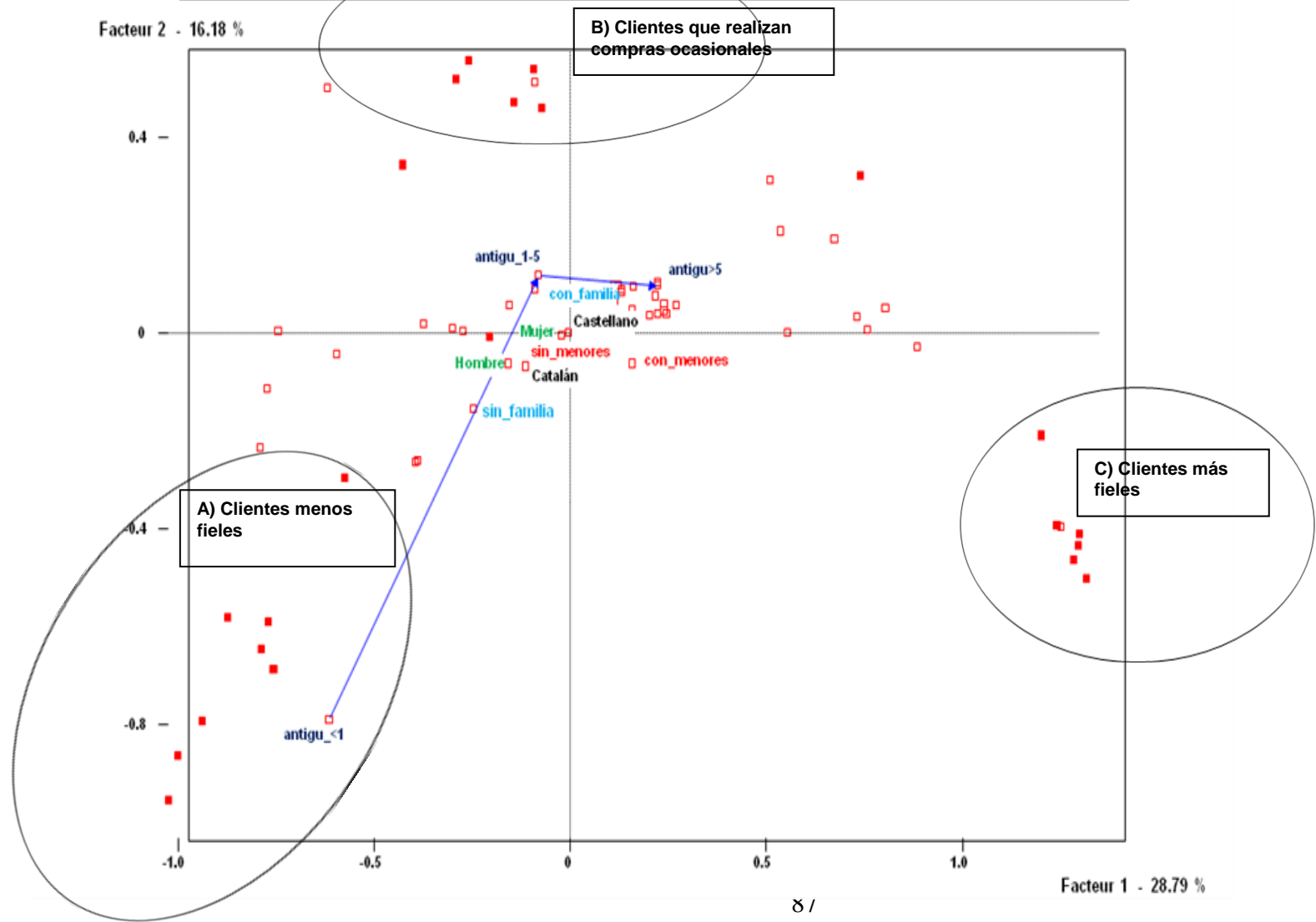


Figura 8.3.1.5: Proyección de las variables suplementarias sociodemográficas (edad)

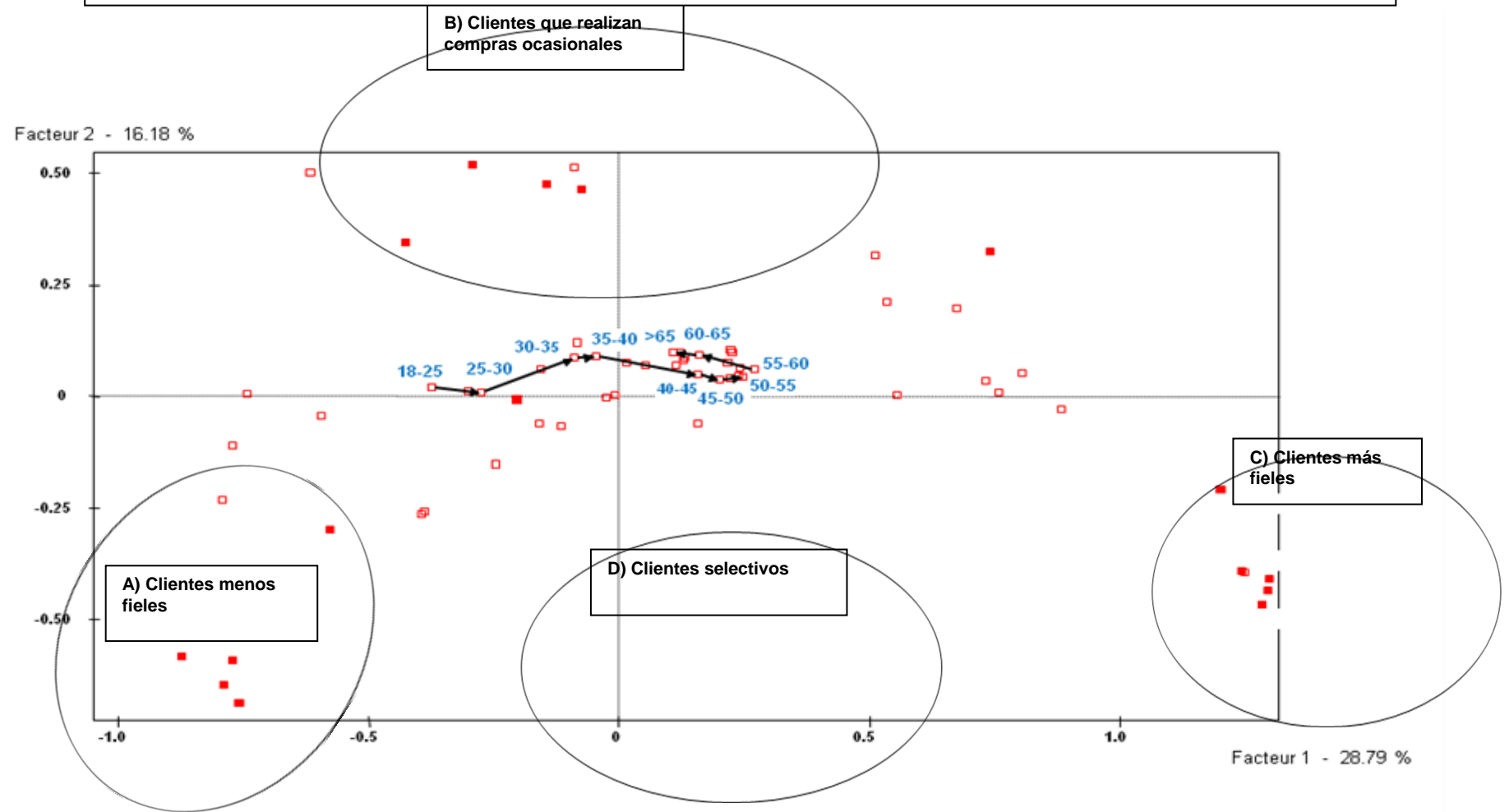
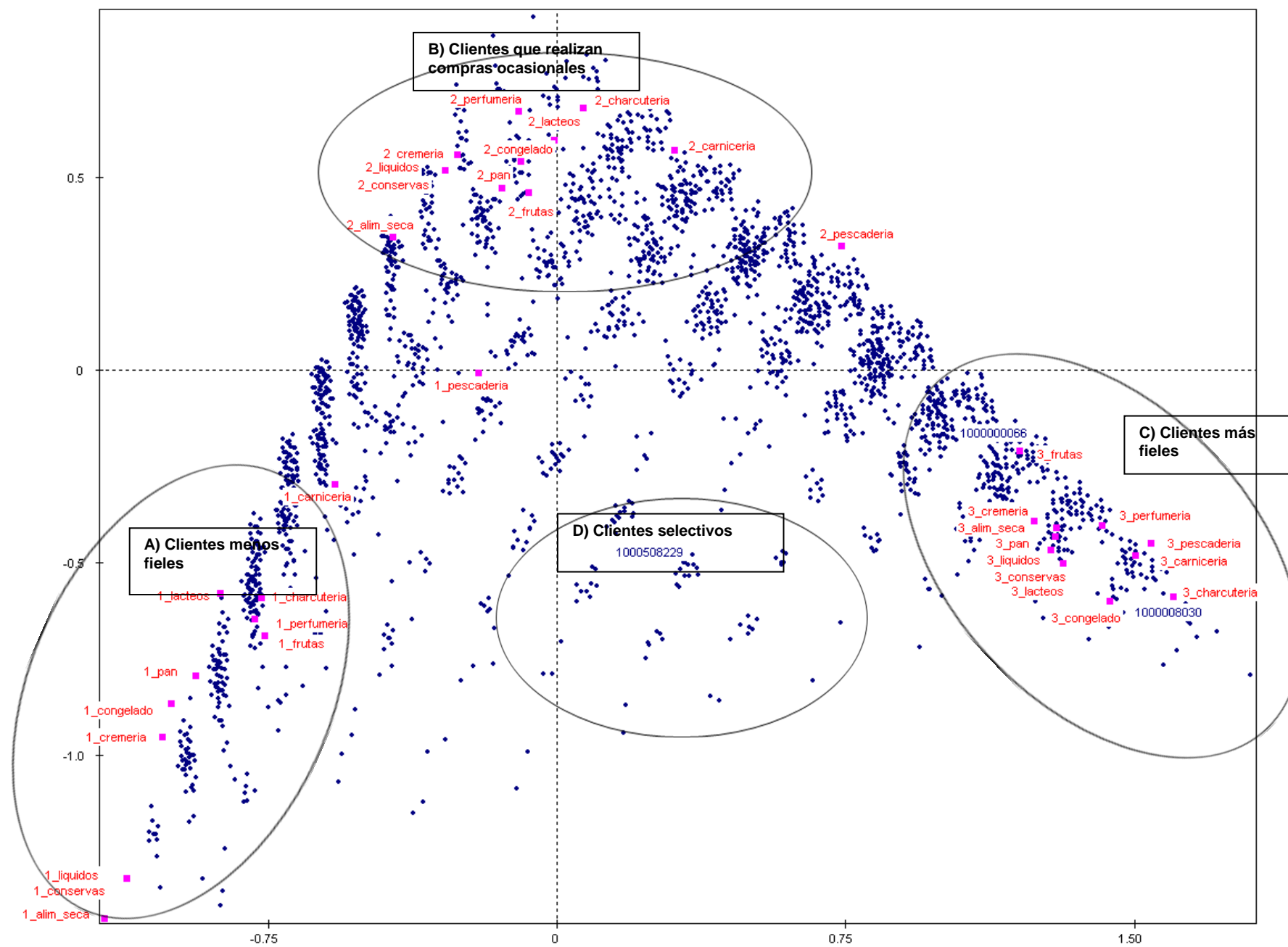


Figura 8.3.1.6 : Proyección de la nube de puntos de los individuos junto con las variables activas.

Facteur 2 - 16.18 %



Facteur 1 - 28.79 %

8.3.2 RESULTADOS SEGMENTACIÓN “SHARE OF WALLET”

Después de haber analizado el Análisis de Correspondencias Múltiples, en este apartado se presentan los resultados de la segmentación tras aplicar el algoritmo de cluster k-Means a los 10 primeros ejes factoriales resultantes del análisis anterior. Para ello se ha utilizado el nodo clustering de SAS MINER.

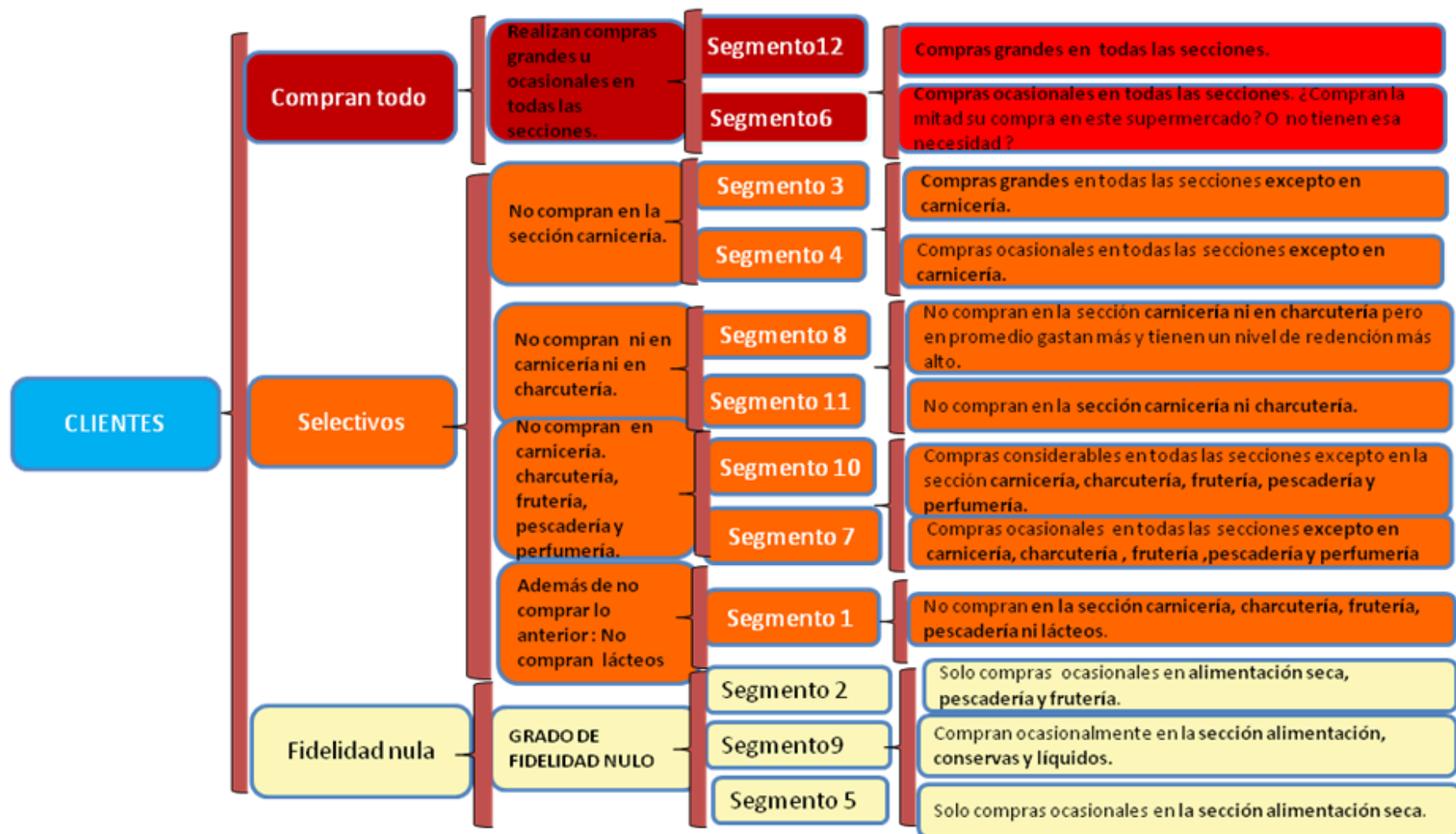
La siguiente figura que se adjunta a continuación (figura 8.3.2.1) muestra de forma esquemática una visión general de los grupos obtenidos. En el esquema se ordenan los doce grupos obtenidos de mayor a menor fidelidad. Más concretamente se ordenan los segmentos de menor a mayor número de secciones con gasto nulo, es decir, aquellos segmentos más fieles son aquellos que cuentan con un número reducido de secciones con gasto nulo mientras aquellos grupos que tienen un número elevado de secciones con gasto nulo se caracterizan por tener un nivel de fidelidad más reducido.

Como se aprecia en la figura 8.3.2.1 se distinguen tres tipos de clientes:

- 1) Los que compran todo en el supermercado.
- 2) Clientes selectivos que no compran en todas las secciones del supermercado.
- 3) Clientes que tienen un nivel de fidelidad muy reducido.

Dentro del grupo que tiende a comprar la mayor parte de su compra en el supermercado distinguimos a su vez otros dos segmentos: el segmento 12 caracterizado por ser el segmento más fiel y por realizar toda su compra en el mismo establecimiento y el conglomerado 6 caracterizado por realizar compras ocasionales en todas las secciones. Dentro de este grupo 6 cabe la duda de saber si es un segmento que está comprando la mitad de su compra en este supermercado o si realmente es un segmento que no tiene esa necesidad de consumo tan fuerte que tiene el segmento 12 y por lo tanto es un segmento realmente fiel.

Figura 8.3.2.1: Esquema segmentación del Share of Wallet.



Dentro del grupo de segmentos denominado como clientes selectivos tenemos a 7 segmentos que son agrupados en 4 subgrupos:

- 1) Clientes que no compran en la sección carnicería.
- 2) Clientes que no compran ni en la sección carnicería ni en charcutería.
- 3) Clientes que no compran en la sección carnicería, charcutería, frutería, pescadería y perfumería.
- 4) Clientes que, además de no comprar en las secciones anteriores, no realizan gasto alguno en la sección lácteos.

Dentro del primer subgrupo (clientes que no compran en la sección carnicería) distinguimos claramente dos tipos de segmentos: un segmento caracterizado por realizar compras grandes (segmento 3) y otro segmento que realiza compras ocasionales (segmento 4). Lógicamente en términos comparativos el segmento 3 tiene un gasto mensual y un número de visitas muy superior al segmento 4. Esta comparación se puede ver observando la figura 8.3.2.2 en la que se representa el gasto mensual medio y el número de visitas medio para cada segmento siendo el tamaño de cada círculo proporcional al porcentaje de clientes que forman parte de cada grupo.

Para los clientes que no compran ni en la sección carnicería ni en charcutería se observan dos grupos diferenciados en términos de gasto mensual y redención. Destaca el grupo 8 que cuenta con un nivel de gasto mensual y un nivel de redención sensiblemente superior al segmento 11. Esta diferencia se ve claramente observando el gráfico 8.3.2.2.

Atendiendo al subgrupo 3 (clientes que no compran en la sección carnicería, charcutería, frutería, pescadería y perfumería) distinguimos dos segmentos: uno caracterizado en realizar compras grandes excepto en las citadas secciones (segmento 10) y otro, segmento 7, que realiza compras ocasionales.

Respecto al último grupo de segmentos caracterizados por un nivel de fidelidad muy reducido aparecen 3 subgrupos que tienen en común que todos ellos al menos realizan compras ocasionales en la sección alimentación seca.

A nivel general, se aprecia una tendencia por parte de los clientes de no comprar productos perecederos en el supermercado. Parece, a la vista de los resultados, que las familias prefieren comprar la carne y el pescado en otros establecimientos como pueden ser los mercados. Se podría preguntar al consumidor para indagar los motivos de que los clientes no compren en las secciones carnicería y pescadería. La hipótesis que se plantea es que parece que la calidad de la carne y el pescado que se ofrece en un supermercado no están tan bien valorados como lo puede estar un mercado especializado.

En el siguiente apartado se procederá a explicar en detalle las características de cada segmento.

Figura 8.3.2.2: Representación de segmentos en función de su gasto y número de visitas.

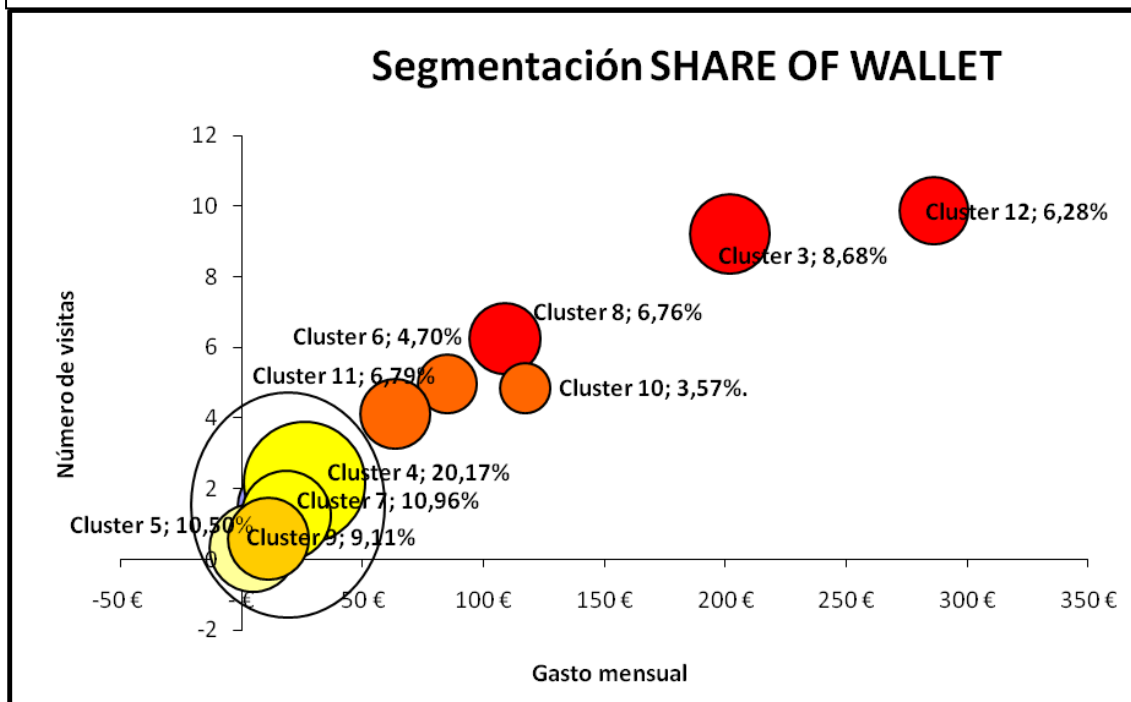
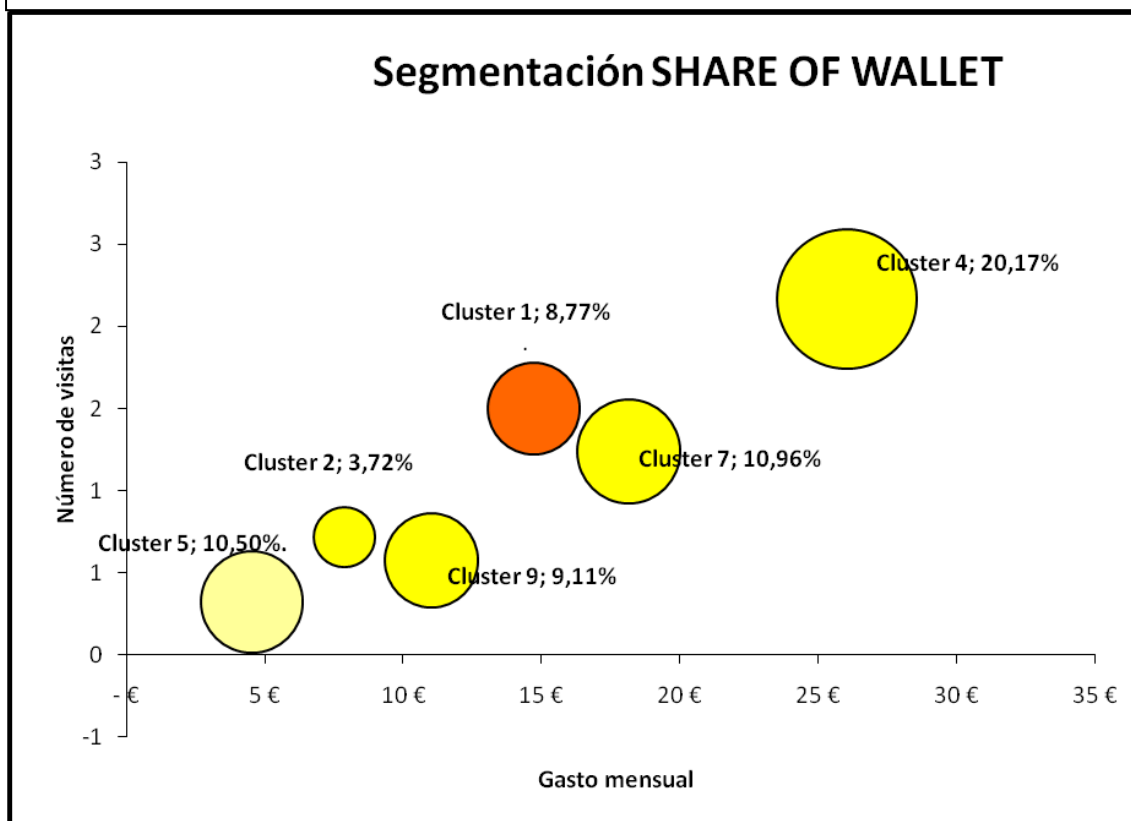


Figura 8.3.2.3: Ampliación de segmentos menos fieles.



8.3.2.1 DISTRIBUCIÓN GENERAL SEGMENTACIÓN “SHARE OF WALLET”

En este apartado se presenta la distribución global de las variables de segmentación (gasto, visitas, compras grandes y redención) y las variables de carácter sociodemográficas para todos los clientes con el objetivo de poder comparar esta distribución general con la distribución de las variables de cada segmento.

Los gráficos presentados son diferentes a los del apartado 5.1 debido sobre todo a que en esta segmentación no hemos eliminado a aquellos clientes que cuentan con un gasto mensual superior a 1.000€ y aquellos que su número de visitas mensuales es mayor que 10. En esta segmentación no se ha realizado ningún filtro por lo que los resultados obtenidos en este dataset proporcionan globalmente unos resultados más altos en términos de fidelidad. Así por ejemplo en los gráficos presentados en el apartado 5.1 el porcentaje de aquellos que gastan más de 200€ al mes es del 4,82% mientras que en este dataset equivale al 7,07%.

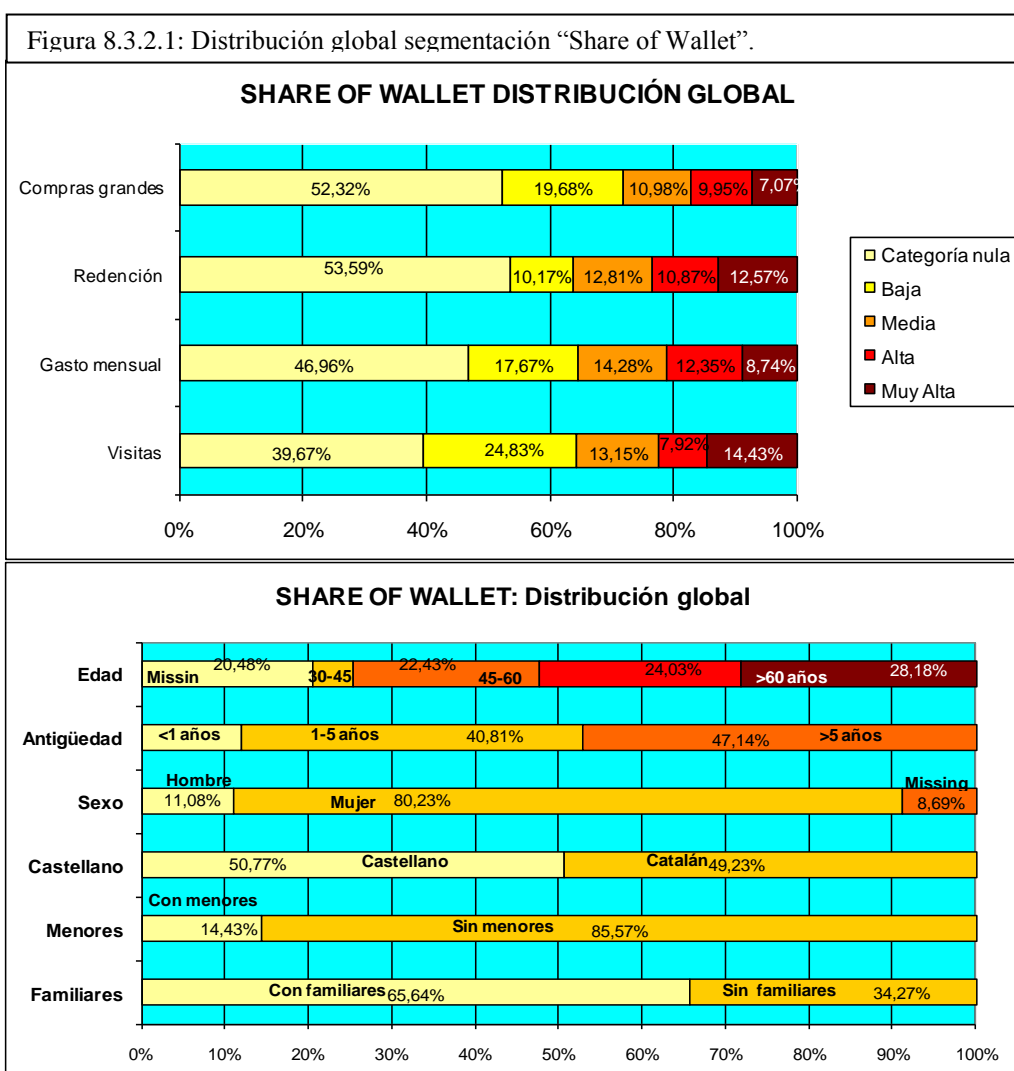


Figura 8.3.2.2. Resumen segmentos obtenidos segmentación en Share of Wallet (1).

**Compran todo
en el
supermercado**

SEGMENTO 12 (6.28%)

- Segmento que tiende a realizar todas las compras en el supermercado.
- Es el segmento más fiel: el que más gasta al mes, el que más veces acude al supermercado y el que más gasta por visita realizada.
- Es el segmento con mayor antigüedad en el supermercado.
- Edad comprendida entre los 45 y 60 años.
- Es el segmento con mayor porcentaje de castellanohablantes (63%).
- Es un segmento que vive con familiares en el hogar.

SEGMENTO 6 (4.70%)

- Se caracteriza por realizar compras ocasionales en todas las secciones.
- Gasto mensual y número de visitas medio.
- Nivel de sensibilidad al cupón elevada.
- Es el segundo segmento con mayor edad caracterizada por la modalidad de **mayores de 60 años**.
- Mayoría de castellanos parlantes.
- Vive con familiares en el hogar.

**No compran en
la sección
carnicería**

SEGMENTO 3 (8.68%)

- Realiza un gasto considerable en todas las secciones excepto en la sección carnicería.
- Es el segundo grupo con un mayor nivel de gasto mensual y número de visitas.
- Nivel de redención a los cupones elevado.
- Predominio de los castellanoparlantes.
- Edad comprendida entre los 45-60 años.
- Tienen una antigüedad superior a los 5 años en el establecimiento.
- Se caracterizan por tener un porcentaje elevado de menores de edad y de familiares en el hogar.

SEGMENTO 4 (20.17%)

- Realiza compras ocasionales en todas las secciones excepto en carnicería.
- Tiene en promedio un gasto mensual y un número reducido de visitas.
- Nivel de redención al cupón medio.
- Tienen una edad similar a la distribución global.
- Se caracterizan por tener una antigüedad inferior a los 5 años
- No hay diferencias entre castellanohablantes y catalanoparlantes.
- Viven con familiares en el hogar.

**No compran ni
en carnicería ni
en charcutería**

SEGMENTO 8 (6.76%)

- No compran en la sección carnicería ni en charcutería pero en promedio gastan más y tienen un nivel de redención más alto.
- Gasto mensual medio-alto.
- Número de visitas alto comparado con su nivel de gasto.
- Nivel de redención elevado.
- Predominio de mayores de 60 años.
- Ligero predominio de clientes con una antigüedad superior a los 5 años.
- Ligero dominio de castellano-parlantes y de clientes que viven con familiares en el hogar.

SEGMENTO 11 (6.79%)

- No compran en la sección carnicería ni en charcutería
- Gasto mensual y número de visitas medio.
- Nivel de sensibilidad al cupón medio.
- Respecto a la edad existe una ligera mayoría de mayores de 60 años.
- La mitad de los clientes de este segmento tienen una antigüedad superior a los 5 años y la otra mitad tiene una antigüedad de 1-5 años.
- Ligera mayoría de castellanoparlantes.
- Es un conglomerado caracterizado por vivir con familiares.

Figura 8.3.2.2. Resumen segmentos obtenidos segmentación en Share of Wallet(2).

No compran en
carnicería.
Charcutería,
frutería,
pescadería y
perfumería.

SEGMENTO 10 (3.57%)

Compras grandes excepto en carnicería, charcutería, frutería, pescadería y perfumería.

- Gasto mensual y número de visitas medio-alto.
- Nivel medio de sensibilidad al cupón.
- Segmento más joven caracterizado por tener una edad comprendida entre 30-45 años.
- Se trata de clientes nuevos que tienen una antigüedad inferior al año.
- Ligero predominio de los catalanoparlantes.
- Viven con familiares.
- Es el segmento en el que más porcentaje de hombres hay.

SEGMENTO 7 (10.96%)

Compras ocasionales excepto en pescadería, carnicería, charcutería, frutería y perfumería

- Gasto mensual, número de visitas y nivel de redención reducido.
- No existen grandes diferencias respecto al idioma, siendo el porcentaje de catalanoparlantes ligeramente superior al de castellanoparlantes.

No compran
lácteos.

SEGMENTO 1 (8.77%)

No compran en la sección carnicería, charcutería, frutería, pescadería ni lácteos.

- Gasto mensual y número de visitas reducido.
- Sensibilidad al cupón media.
- Grupo de una edad elevada, siendo el grupo de mayores de 60 años el más predominante.
- No vive con menores de edad en el hogar
- Antigüedad inferior a 5 años.

Grado de
fidelidad
reducido

SEGMENTO 2 (3,77%)

Solo compran ocasionalmente en la sección alimentación seca, pescadería y frutería.

- Gasto mensual, número de visitas muy reducidas y nivel de redención muy reducido.
- Dentro del subgrupo de los menos fieles, este segmento es el de mayor edad y el de mayor antigüedad.
- Se caracteriza por no vivir con menores de edad y el porcentaje de clientes que viven sin familiares es considerable (33%).

SEGMENTO 9 (9,11%)

Solo compran ocasionalmente en la sección alimentación seca, conservas y líquidos.

- Gasto mensual, número de visitas y nivel de redención muy reducido.
- Dentro del subgrupo de los menos fieles, este segmento es el más joven. Existe un predominio del grupo de 30-45 años
- Predominio de catalanoparlantes.**
- No viven con menores de edad y el porcentaje de clientes que viven sin familiares es considerable.

SEGMENTO 5 (10,50%)

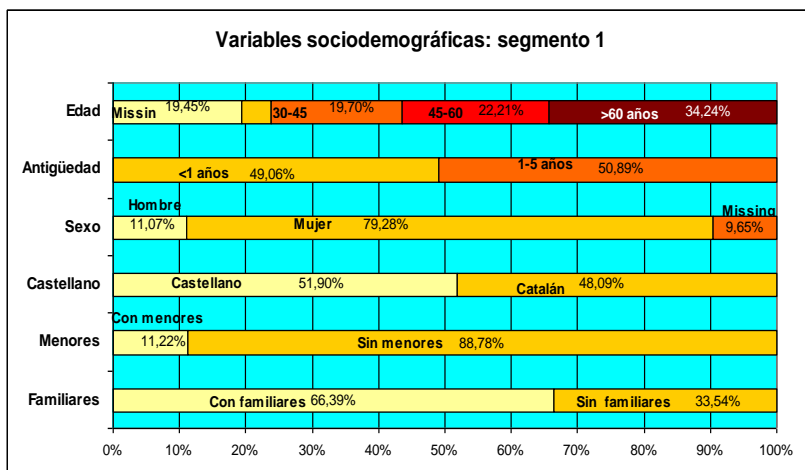
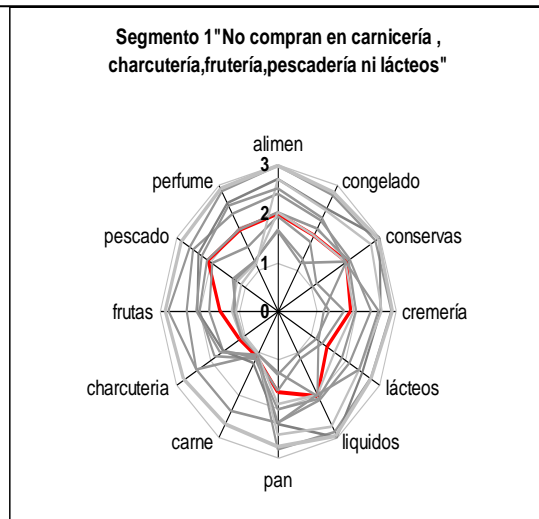
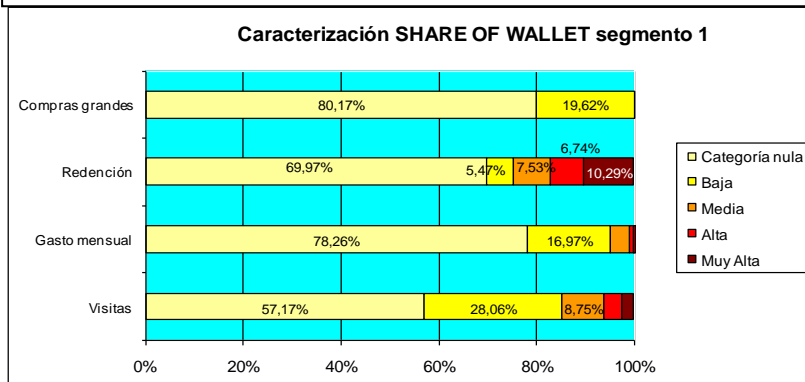
Solo compran ocasionalmente en la sección alimentación seca.

- Gasto mensual, número de visitas y nivel de redención muy reducido.
- Dentro del subgrupo es el que menos gasta: gasto inferior a 5€.
- Predominio de catalanoparlantes.**
- No viven con menores de edad y el porcentaje de clientes que viven sin familiares es considerable.

8.3.2.2.DESCRIPCIÓN SEGMENTO 1 : "No compran en la sección carnicería, charcutería, frutería, pescadería ni lácteos.

Tamaño: 8.77%

8.3.2.2. Caracterización segmentación SHARE OF WALLET segmento 1



	Segmento1	Global
% redención	2,96%	3,70%
Gasto mensual	14,71 €	65,60 €
Visitas mensual	1,50	3,35
Gasto por visita	15,26 €	20,78 €
Num.grandes categorías	0,27	2,45
Núm.Menores	1,43	1,45
Antigüedad	4,30	3,94
Num familiares	1,95	1,98
Edad	57,61	54,91

En la parte izquierda se caracteriza al segmento 1 mostrando la distribución de las variables de fidelidad y socio-demográficas. En la derecha se muestra un gráfico de "tela de araña". En él se muestra por cada segmento como distribuye su gasto mensual cada grupo entre las diferentes secciones. La variable gasto mensual ha sido categorizada tal como se ha explicado en el apartado 2.6.2 y en el gráfico 2.6.2.2

En este caso el comportamiento del segmento 1 se representa con una línea de color rojo. En este caso se aprecia que el segmento 1 se caracteriza porque sus clientes tienden a no realizar compras considerables en carnicería, pescadería, charcutería, frutería, pescadería y lácteos

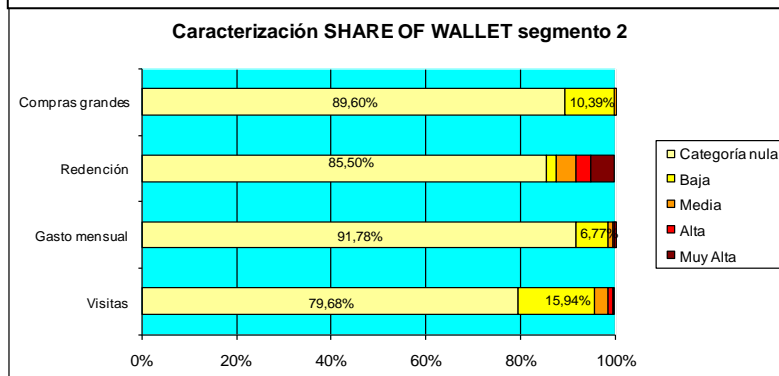
Este segmento destaca sobre todo porque no compran en la sección lácteos. Además este grupo representa el 8,77% de los clientes totales. Sus características se exponen a continuación:

- Gasto mensual y número de visitas reducido.
- Sensibilidad al cupón media.
- Respecto a la edad hay que decir que es un grupo de una edad elevada, siendo el grupo de mayores de 60 años el más predominante.
- Es un grupo caracterizado por no tener menores de edad en sus hogares.
- Tienen una antigüedad inferior a 5 años.

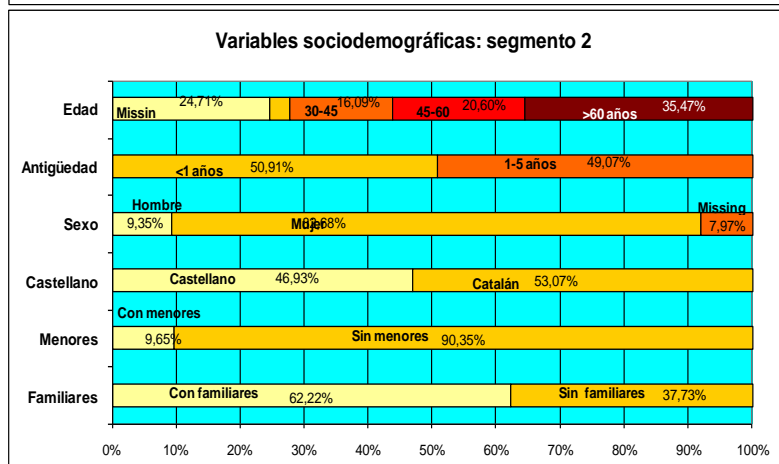
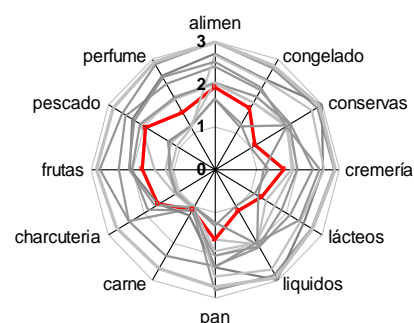
8.3.2.3 DESCRIPCIÓN SEGMENTO 2 : "Solo compran ocasionalmente en la sección alimentación seca, pescadería y frutería".

Tamaño: 3.77%

8.3.2.3. Caracterización segmentación SHARE OF WALLET segmento 2



Segmento 2 "Solo compras ocasionales en alimentación seca, pescadería y frutería"



	Segmento2	Global
% redención	1,53%	3,70%
Gasto mensual	7,88 €	65,60 €
Visitas mensual	0,72	3,35
Gasto por visita	17,21 €	20,78 €
Num.grandes categorías	0,12	2,45
Núm.Menores	1,43	1,45
Antigüedad	4,16	3,94
Num familiares	1,82	1,98
Edad	60,57	54,91

Este segmento ha sido clasificado dentro del subgrupo de segmentos considerados con un grado de fidelidad nulo. Este segmento representa el 3,77% de la totalidad de los clientes,

Sus características son las siguientes:

- Gasto mensual, número de visitas muy reducidas y nivel de redención muy reducido.

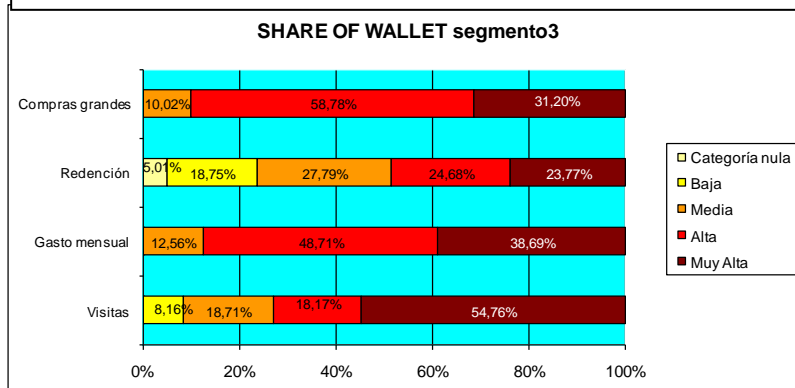
- Dentro del subgrupo de los menos fieles, este segmento es el de mayor edad y el de mayor antigüedad.

- Se caracteriza por no vivir con menores de edad y el porcentaje de clientes que viven sin familiares es considerable (33%).

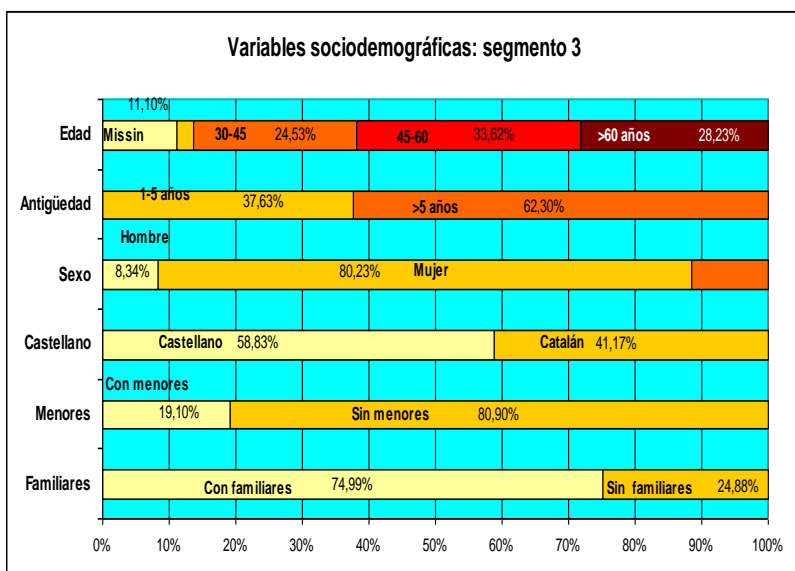
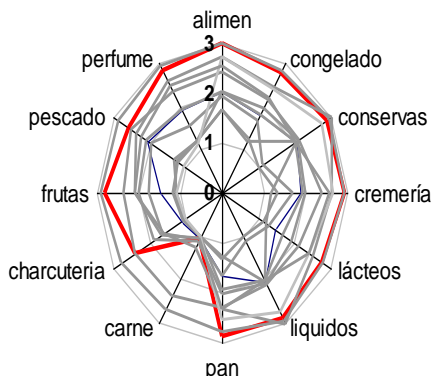
8.3.2.4 DESCRIPCIÓN SEGMENTO 3: “Compran todo en el supermercado excepto en la sección carnicería”

Tamaño: 8.68%

8.3.2.4. Caracterización segmentación SHARE OF WALLET segmento 3



Segmento 3 "Compras grandes en todas las secciones excepto en carnicería"



	Segmento3	Global
% redención	7,03%	3,70%
Gasto mensual	201,59 €	65,60 €
Visitas mensual	9,22	3,35
Gasto por visita	28,48 €	20,78 €
Num.grandes categorías	8,60	2,45
Núm.Menores	1,46	1,45
Antigüedad	5,05	3,94
Num familiares	2,37	1,98
Edad	53,78	54,91

Este segmento representa el 8,68% de todos los clientes.

Lo más significativo de este segmento de clientes es que realiza un gasto considerable en todas las secciones excepto en la sección carnicería.

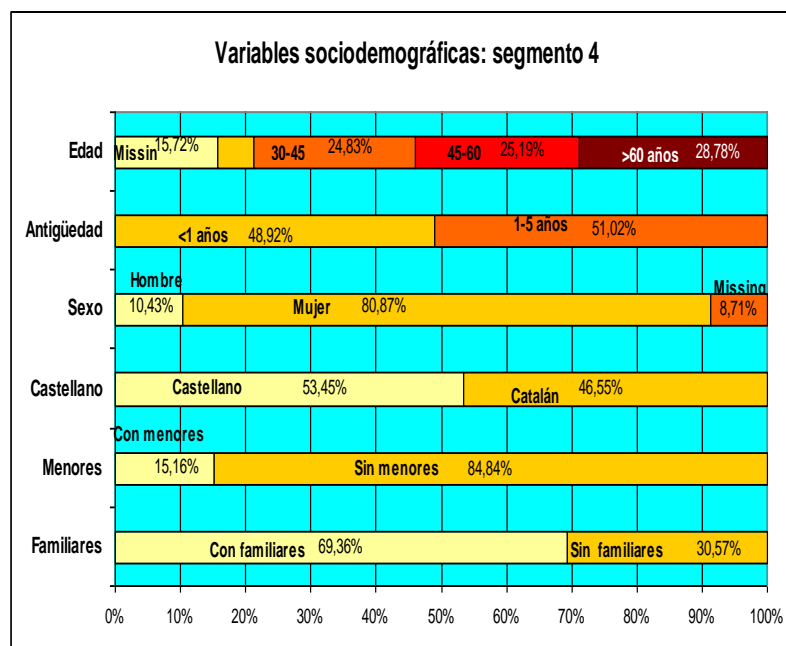
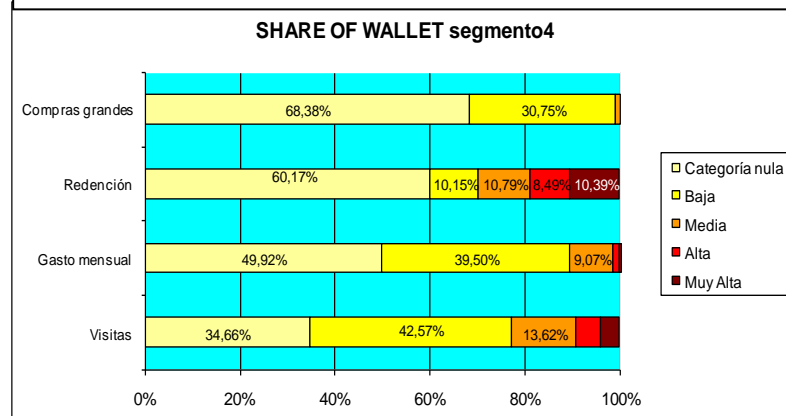
Las características de este segmento son las siguientes:

- Es el segundo grupo con un mayor nivel de gasto mensual y número de visitas.
- Tienen un nivel de redención a los cupones bastante elevado.
- Existe un predominio de castellanoparlantes.
- El grupo de edad más influyente es el comprendido entre los 45-60 años.
- Tienen una antigüedad superior a los 5 años en el establecimiento.
- Se caracterizan por tener un porcentaje elevado de menores de edad y de familiares en el hogar.

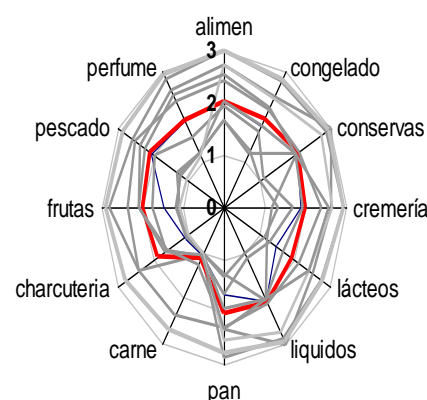
8.3.2.5 ESCRIPCIÓN SEGMENTO 4: “Compras ocasionales en todas las secciones excepto en carnicería”

Tamaño: 20.17%

8.3.2.5. Caracterización segmentación SHARE OF WALLET segmento 4



Segmento 4 "Compras ocasionales excepto en carnicería"



	Segmento4	Global
% redención	3,08%	3,70%
Gasto mensual	26,05 €	65,60 €
Visitas mensual	2,17	3,35
Gasto por visita	17,97 €	20,78 €
Num.grandes categorías	0,50	2,45
Núm.Menores	1,40	1,45
Antigüedad	4,39	3,94
Num familiares	2,07	1,98
Edad	54,71	54,91

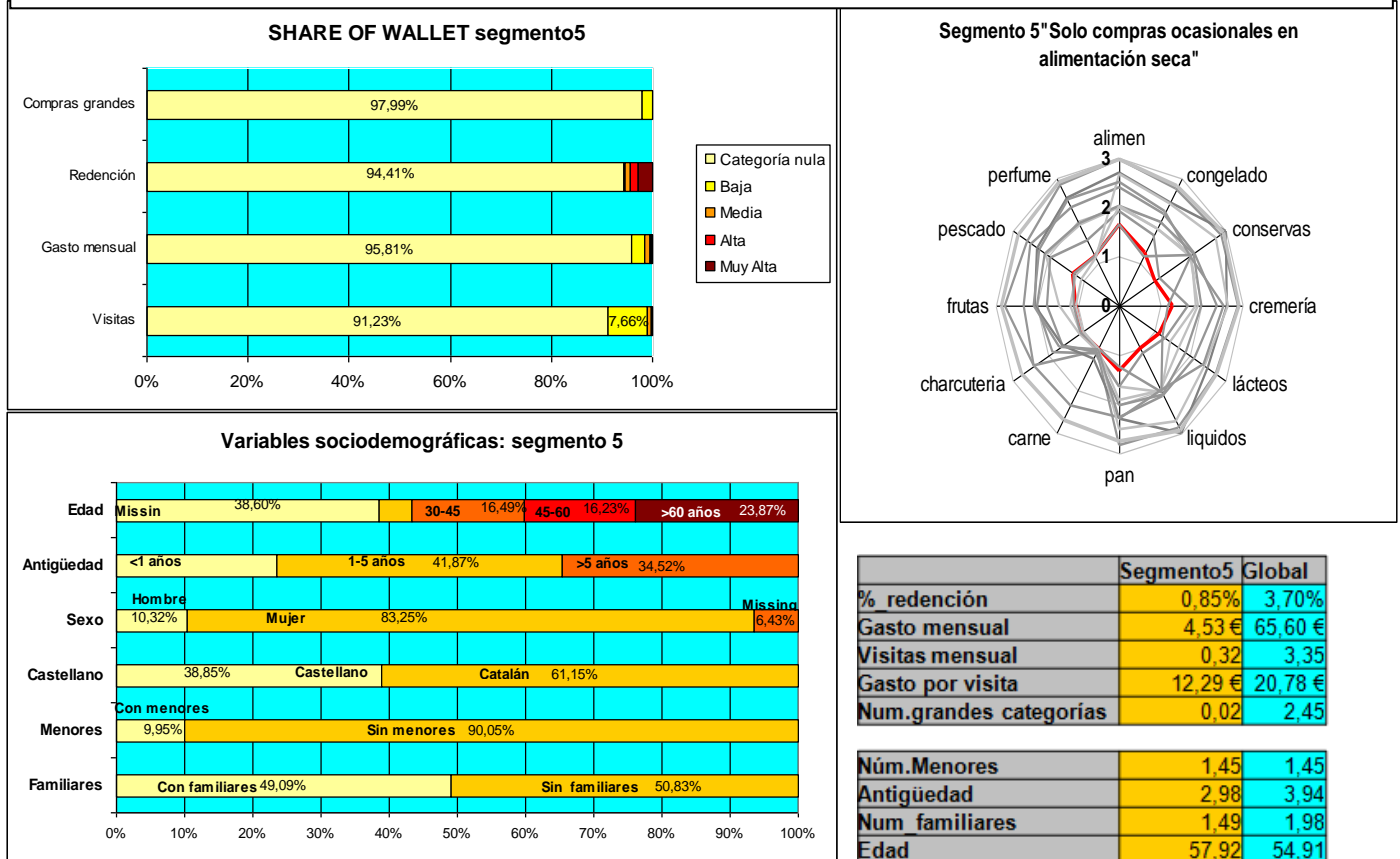
Este segmento agrupa al 20,17% de todos los clientes y las características principales de este segmento son las siguientes:

- Se caracteriza por tener en promedio un gasto mensual y un número reducido de visitas.
- Tienen un nivel medio de redención al cupón.
- Tienen una edad similar a la distribución global. En este segmento no existe el predominio del grupo de edad de los 45-60 años.
- Se caracterizan por tener una antigüedad inferior a los 5 años. La mitad del segmento cuenta con una antigüedad inferior al año mientras que la otra mitad cuenta con una antigüedad comprendida entre 1 y 5 años.
- No hay diferencias entre castellanohablantes y catalanoparlantes.
- Se caracterizan por vivir con familiares en el hogar.

8.3.2.6 DESCRIPCIÓN SEGMENTO 5 : "Solo compran ocasionalmente en la sección alimentación seca"

Tamaño: 10.5%

8.3.2.6. Caracterización segmentación SHARE OF WALLET segmento 5



Este segmento con fidelidad nula representa el 10,50% de la totalidad de los clientes,

Sus características son las siguientes:

- Gasto mensual, número de visitas muy reducidas y nivel de redención muy reducido. Dentro del subgrupo de escasa fidelidad es el que menos gasta: gasto inferior a 5€.

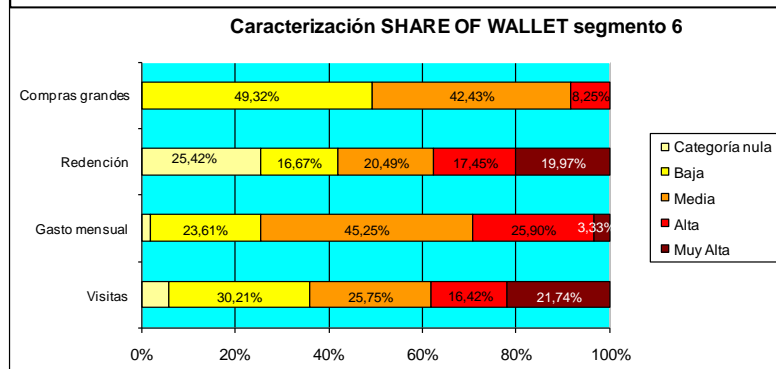
- Existe un predominio de catalanoparlantes

- Se caracteriza por no vivir con menores de edad y el porcentaje de clientes que viven sin familiares es considerable (49,09%).

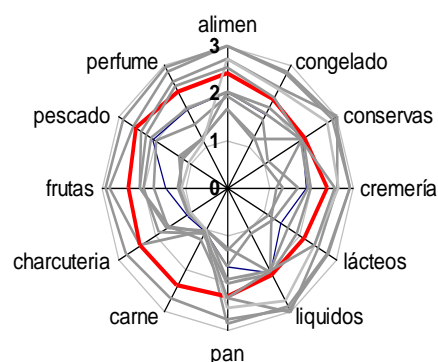
8.3.2.7 DESCRIPCIÓN SEGMENTO 6: "Realizan compras ocasionales en todas las secciones".

Tamaño: 4.70%

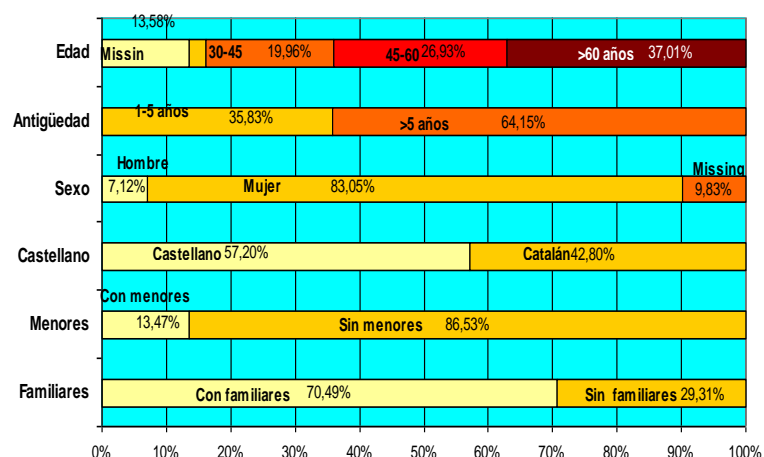
8.3.2.7. Caracterización segmentación SHARE OF WALLET segmento 6



Segmento 6 "Compras ocasionales en todas las secciones"



Variables sociodemográficas: segmento 6



	Segmento6	Global
% redención	5,64%	3,70%
Gasto mensual	85,10 €	65,60 €
Visitas mensual	4,97	3,35
Gasto por visita	23,86 €	20,78 €
Num.grandes categorías	3,63	2,45

Núm.Menores	1,46	1,45
Antigüedad	5,08	3,94
Num familiares	2,16	1,98
Edad	57,02	54,91

Como hemos expresado en el título de este apartado y tal como se muestra en el gráfico de "tela de araña" este segmento se caracteriza por realizar compras ocasionales en todas las secciones.

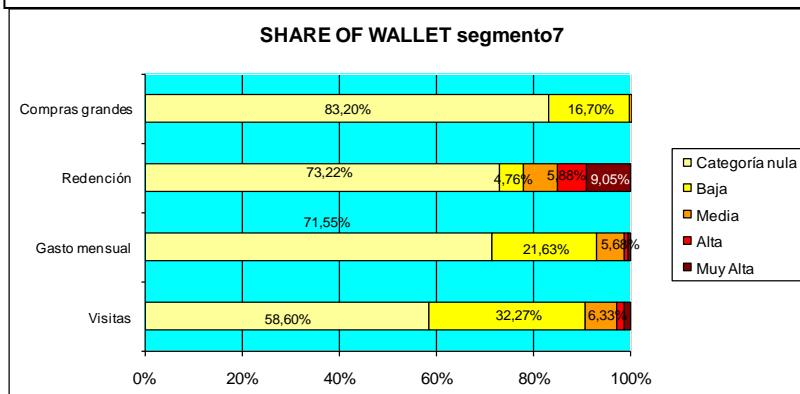
El segmento 6, que representa el 4,70% de todos los clientes, se caracteriza por lo siguiente:

- Tienen un gasto mensual y un número de visitas que se puede considerar como medio.
- Cuentan con un nivel de sensibilidad al cupón elevada.
- Es el segundo segmento con mayor edad caracterizada por la modalidad de mayores de 60 años.
- Existe una mayoría de castellanoparlantes.
- Vive con familiares en el hogar.
- Cuenta con un porcentaje de menores de edad similar a la media.

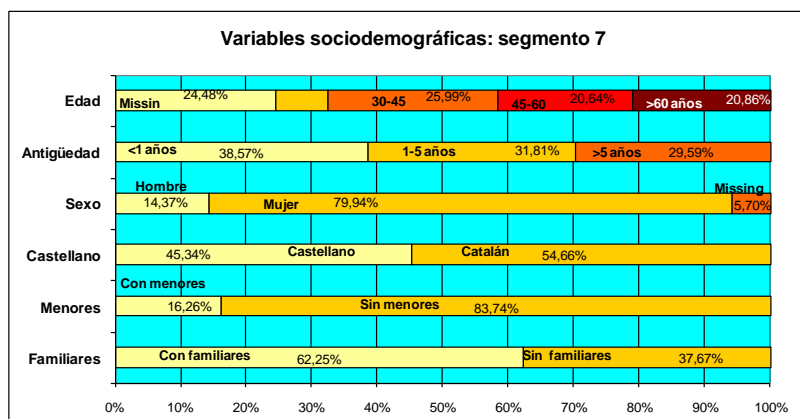
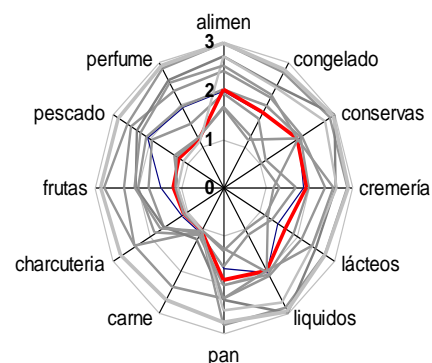
8.3.2.8 DESCRIPCIÓN SEGMENTO 7: “Compras ocasionales excepto en pescadería, carnicería ,charcutería , frutería y perfumería”

Tamaño: 10.96%

8.3.2.8. Caracterización segmentación SHARE OF WALLET segmento 7



Segmento 7" No compran en carnicería , charcutería,frutería,pescadería y perfumería"



	Segmento7	Global
% redención	2,61%	3,70%
Gasto mensual	18,16 €	65,60 €
Visitas mensual	1,24	3,35
Gasto por visita	20,79 €	20,78 €
Num.grandes categorías	0,23	2,45

Núm.Menores	1,46	1,45
Antigüedad	2,54	3,94
Num familiares	1,88	1,98
Edad	51,01	54,91

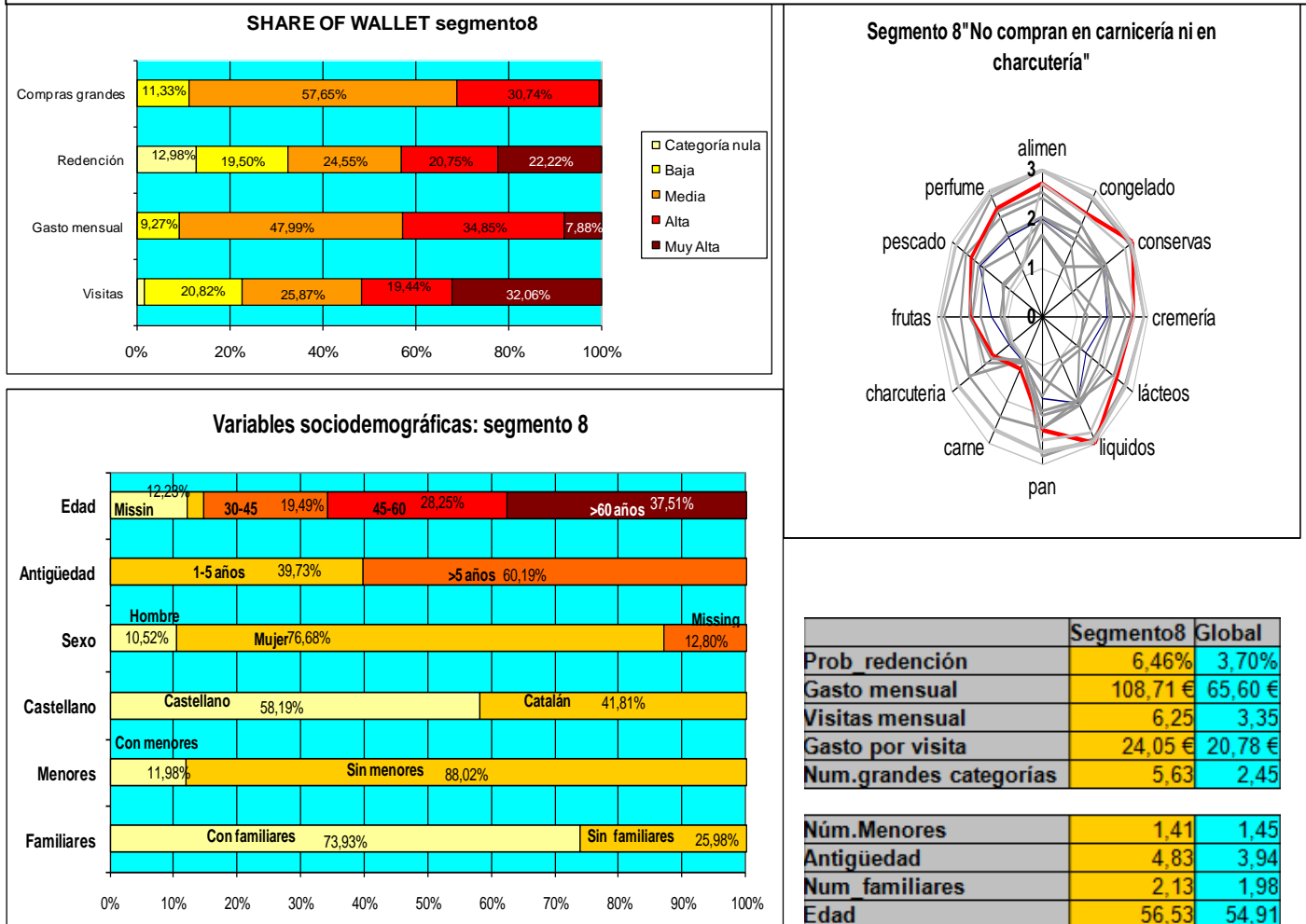
Este segmento agrupa al 10,96% de todos los clientes totales caracterizándose por lo siguiente:

- Gasto mensual, número de visitas y nivel de redención reducido.
- No existen grande diferencias respecto al idioma, siendo el porcentaje de catalanoparlantes ligeramente superior al de castellanoparlantes.

8.3.2.9 DESCRIPCIÓN SEGMENTO 8: "No compran en la sección carnicería ni charcutería pero en promedio gastan más y tienen un nivel de redención más alto"

Tamaño: 6.76%

8.3.2.9. Caracterización segmentación SHARE OF WALLET segmento 8



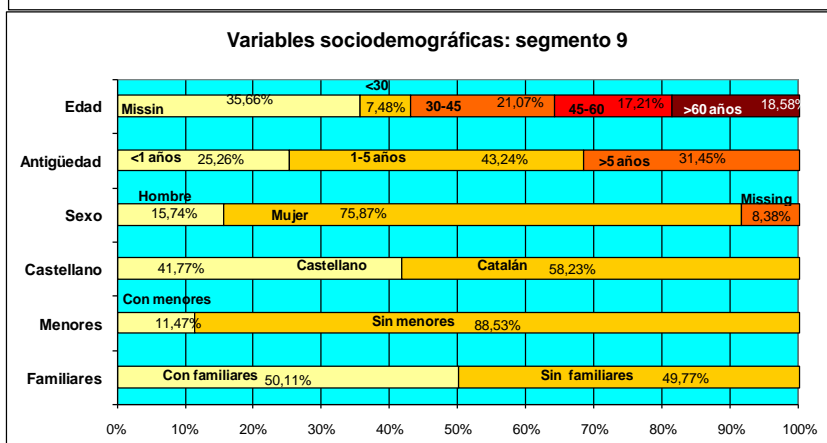
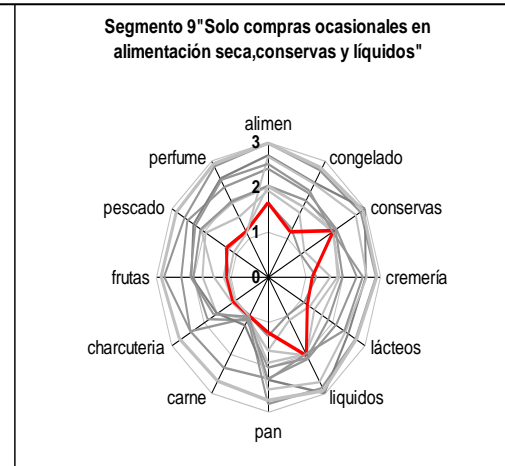
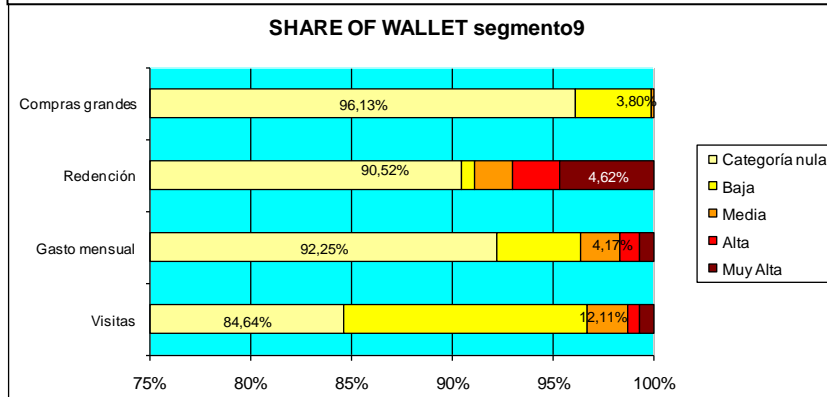
Este segmento agrupa al 6,76 % del total de los clientes y se caracteriza por lo siguiente:

- Gasto mensual medio-alto.
- Número de visitas alto comparado con su nivel de gasto. El segmento 10 que cuenta con un nivel de gasto similar tiene un promedio aproximado de 4 visitas.
- Cuenta con un nivel de redención elevado.
- Respecto a la edad, el grupo predominante es el de mayores de 60 años.
- Existe un ligero predominio de clientes con una antigüedad superior a los 5 años.
- Hay un ligero dominio de castellano-parlantes y de clientes que viven con familiares en el hogar.

8.3.2.10 DESCRIPCIÓN SEGMENTO 9 : "Solo compran ocasionalmente en la sección alimentación seca, conservas y líquidos"

Tamaño: 9.11%

8.3.2.10. Caracterización segmentación SHARE OF WALLET segmento 9



	Segmento9	Global
% redención	1,29%	3,70%
Gasto mensual	11,03 €	65,60 €
Visitas mensual	0,58	3,35
Gasto por visita	15,30 €	20,78 €
Num.grandes categorías	0,07	2,45

Núm.Menores	1,44	1,45
Antigüedad	2,84	3,94
Num familiares	1,56	1,98
Edad	53,29	54,91

Este segmento con fidelidad nula representa el 9,11% de la totalidad de los clientes.

Sus características son las siguientes:

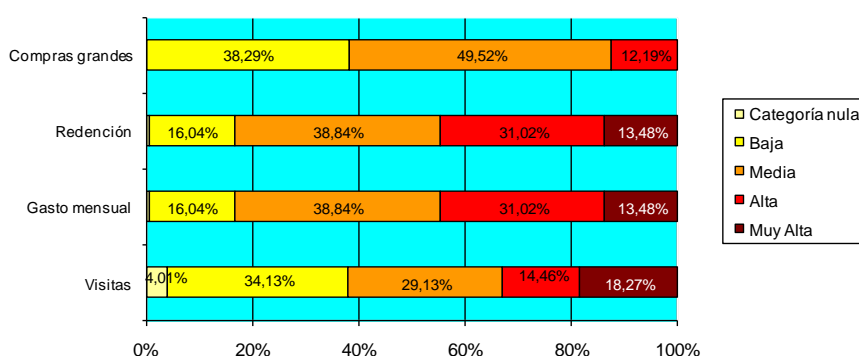
- Gasto mensual, número de visitas y nivel de redención muy reducido.
- Dentro del subgrupo de los menos fieles, este segmento es el más joven. Existe un predominio del grupo de 30-45 años
- Existe un predominio de catalanoparlantes
- Se caracteriza por no vivir con menores de edad y el porcentaje de clientes que viven sin familiares es considerable (49,77%).

8.3.2.11 DESCRIPCIÓN SEGMENTO 10 : "Compras grandes excepto en carnicería, charcutería, frutería, pescadería y perfumería"

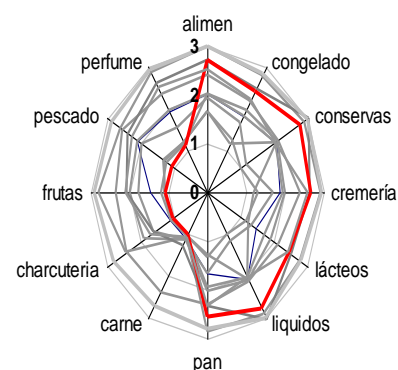
Tamaño: 3.57%

8.3.2.11. Caracterización segmentación SHARE OF WALLET segmento 10

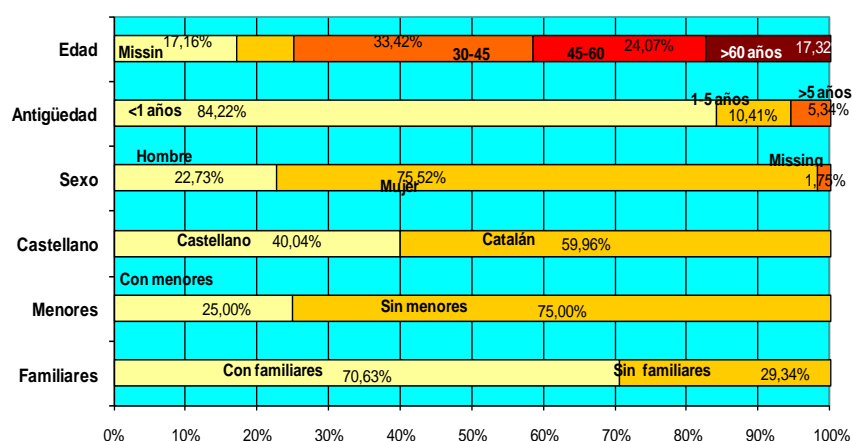
SHARE OF WALLET segmento10



Segmento 10 "No compran en carnicería, charcutería, frutería, pescadería y perfumería"



Variables sociodemográficas: segmento 10



	Segmento10	Global
Prob redención	5,65%	3,70%
Gasto mensual	117,15 €	65,60 €
Visitas mensual	4,85	3,35
Gasto por visita	33,92 €	20,78 €
Num.grandes categorías	4,28	2,45
Núm.Menores	1,54	1,45
Antigüedad	0,51	3,94
Num familiares	2,08	1,98
Edad	47,07	54,91

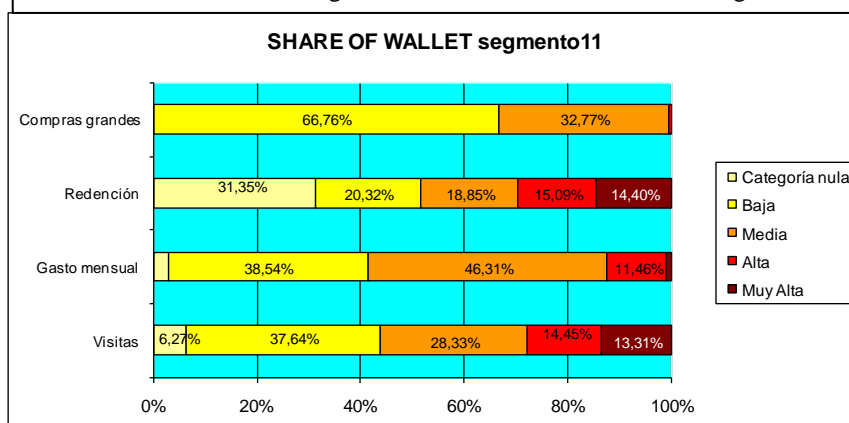
Este segmento 10 aglutina al 3,57% del total de todos los clientes y cuenta con las siguientes características:

- Gasto mensual y número de visitas medio-alto.
- Nivel medio de sensibilidad al cupón.
- Se trata del segmento más joven caracterizado por tener una edad comprendida entre 30-45 años.
- Se trata de clientes nuevos que tienen una antigüedad inferior al año.
- En este grupo existe un ligero predominio de los catalanoparlantes.
- Es un conglomerado caracterizado por vivir con familiares.
- Es el segmento en el que más porcentaje de hombres hay.

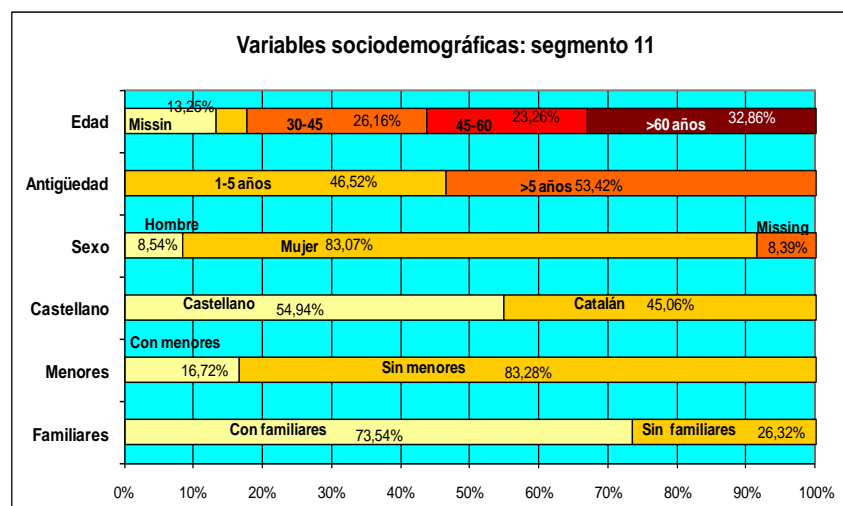
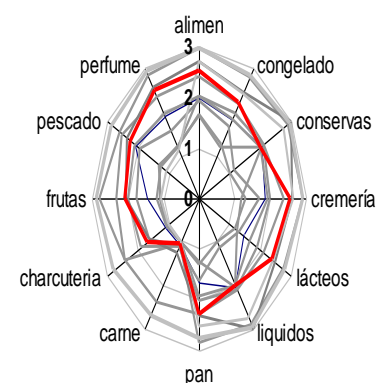
8.3.2.12 DESCRIPCIÓN SEGMENTO 11: "No compran en la sección carnicería ni charcutería"

Tamaño: 6.79%

8.3.2.12. Caracterización segmentación SHARE OF WALLET segmento 11.



Segmento 11 "No compran en carnicería ni en charcutería"



	Segmento11	Global
% redención	4,59%	3,70%
Gasto mensual	63,42 €	65,60 €
Visitas mensual	4,12	3,35
Gasto por visita	21,24 €	20,78 €
Num.grandes categorías	2,90	2,45

Núm.Menores	1,47	1,45
Antigüedad	4,51	3,94
Num familiares	2,14	1,98
Edad	55,45	54,91

Este segmento agrupa al 6,79% del total de los clientes y se caracteriza por:

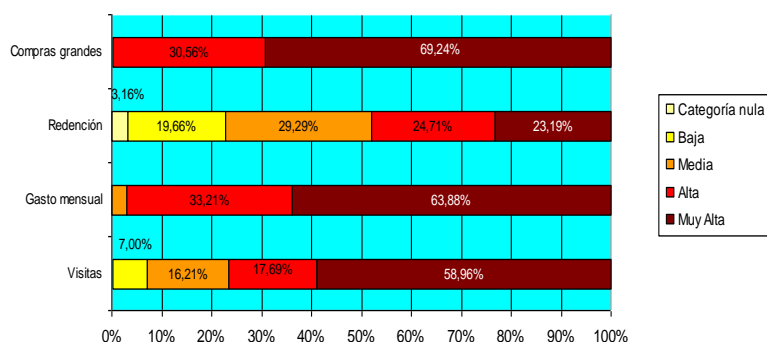
- Gasto mensual y número de visitas medio.
- Nivel de sensibilidad al cupón medio.
- Respecto a la edad existe una ligera mayoría de mayores de 60 años.
- La mitad de los clientes de este segmento tienen una antigüedad superior a los 5 años y la otra mitad tiene una antigüedad de 1-5 años.
- Hay una ligera mayoría de castellanoparlantes.
- Es un conglomerado caracterizado por vivir con familiares.

8.3.2.13 DESCRIPCIÓN SEGMENTO 12: "Los que compran todo en el supermercado"

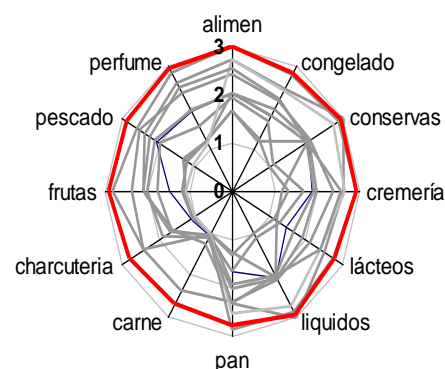
Tamaño: 6.28%

8.3.2.13. Caracterización segmentación SHARE OF WALLET segmento 12.

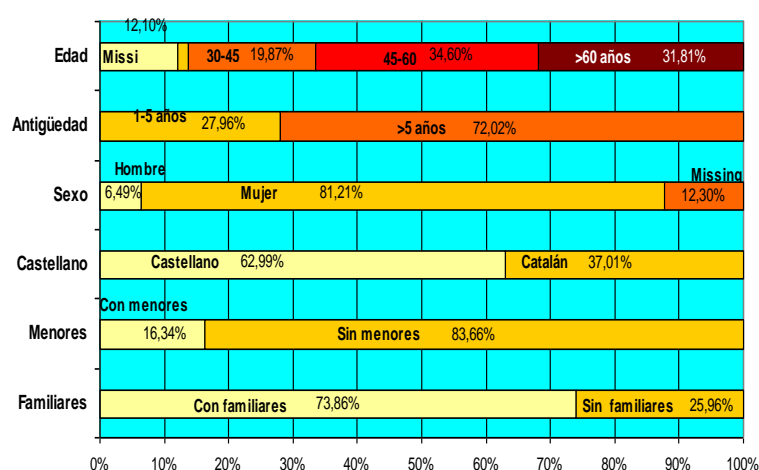
Caracterización SHARE OF WALLET segmento 12



Segmento 12 "Los que compran todo en el supermercado"



Variables sociodemográficas: segmento 12



	Segmento12	Global
% redención	7,15%	3,70%
Gasto mensual	286,16 €	65,60 €
Visitas mensual	9,86	3,35
Gasto por visita	37,28 €	20,78 €
Num.grandes categorías	10,20	2,45

Núm.Menores	1,47	1,45
Antigüedad	5,49	3,94
Num familiares	2,43	1,98
Edad	54,76	54,91

El segmento 12, cuyo tamaño es del 6,28 % del total de los clientes, se caracteriza por lo siguiente:

- Es el segmento más fiel, el que más gasta al mes, el que más veces acude al supermercado y el que más gasto por visita realiza.
- Es el segmento con mayor antigüedad en el supermercado.
- Es el segmento que cuenta con el mayor porcentaje de personas con una edad comprendida entre los 45 y 60 años.
- Es el segmento con mayor porcentaje de castellanoparlante (63%).
- Es un segmento que vive con familiares en el hogar.

9.SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO

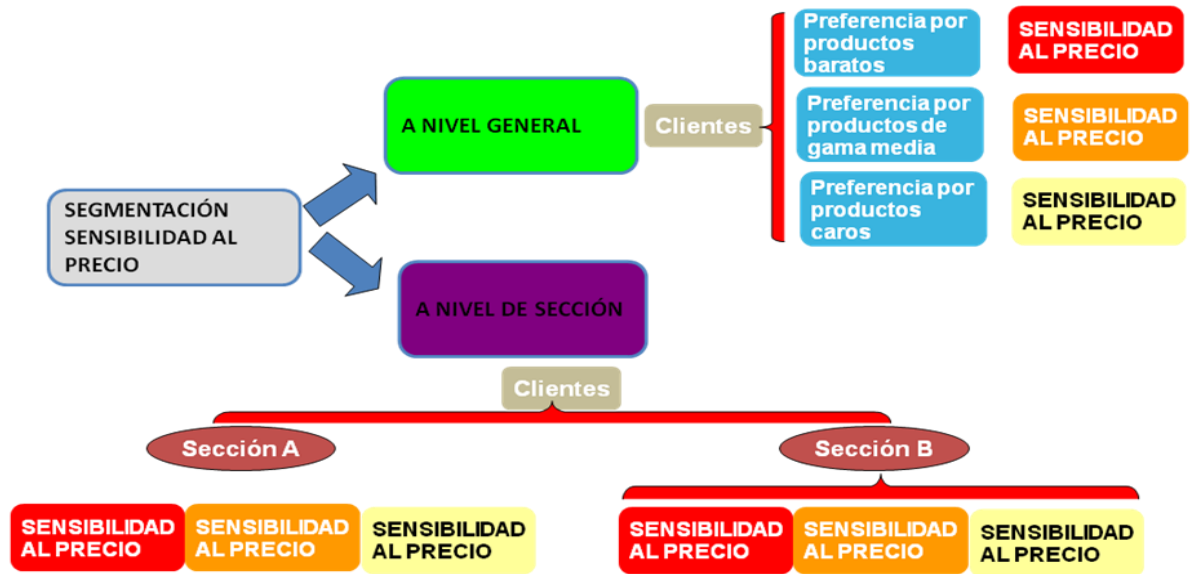
9.1 DESCRIPCIÓN DE LA SEGMENTACIÓN

Como el título indica, en este apartado se segmentará a los clientes en función de su nivel de sensibilidad al precio. Se dice que un cliente es sensible al precio cuando tiende a comprar productos baratos. Un cliente será insensible al precio cuando muestra una preferencia en sus hábitos de compra por productos de precios más elevados.

Tal como se muestra en el esquema 9.1.1 se segmentará a los clientes según su nivel de sensibilidad al precio en dos niveles por lo que se realizará dos segmentaciones:

- 1) Se pretende **segmentar a los clientes en función de su sensibilidad al precio de forma general**. Más concretamente se pretende detectar tres tipologías de clientes:
 - Segmentos de clientes que tienen preferencia por productos baratos, por lo que su nivel de sensibilidad al precio es alta.
 - Segmentos de clientes que tienen preferencia por productos de precio medio, con lo que su nivel de sensibilidad al precio es medio.
 - Segmentos de clientes que tienen preferencia por productos considerados como caros, lo que supone un nivel de sensibilidad al precio reducido.
- 2) Se aplicará una **segmentación en función de la sensibilidad al precio a nivel de sección**, es decir, el objetivo es detectar segmentos de clientes homogéneos que son sensibles al precio en unas secciones pero insensibles en otras. En otras palabras, se quiere detectar grupos de clientes que compran productos caros en unas secciones determinadas pero que compran productos baratos en otras secciones. Por ejemplo, puede existir un segmento de consumidores que compran productos baratos en todas las secciones excepto en la sección perfumería en la que optan por comprar productos caros.

Figura 9.1.1 Esquema segmentación de clientes en función de la sensibilidad al precio.



FILTRADO DE DATOS

Se parte de un conjunto de datos que contienen las compras efectuadas por todos los clientes durante el año 2010 (enero-diciembre 2010).

De esta tabla de datos se eliminan aquellos clientes que han comprado menos de 24 productos en un año. La razón es que puede haber clientes que compren pocos productos y con esa información sea difícil determinar su grado de sensibilidad al precio.

9.2 METODOLOGÍA DE SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO A NIVEL GENERAL

La metodología aplicada en esta segmentación, que está esquematizada en el esquema 9.2.1, se puede desglosar en varias etapas:

1) Observando el comportamiento de compra durante un año se clasifica a los productos comprados en: **productos caros, productos de gama media y productos baratos**. En el apartado 2.5.4 se explicó con detalle el procedimiento de clasificación de los productos.

2) Después se determina el número de productos baratos, caros y de precio medio comprados por cada cliente durante el año 2010.

Con lo anterior se calcula la proporción de productos baratos, de productos de gama media y de productos caros adquiridos por cada cliente, obteniéndose una tabla como la que se muestra a continuación.

Tabla 9.2.1.

COMPRAS AÑO 2010							
Id Cliente	Número de productos comprados	Núm. Productos precio medio	Núm. Productos Baratos	Núm. Productos Caros	Proporción productos precio medio	Proporción productos caros	Proporción productos baratos
1001							
1002							
1003							

3) Finalmente se aplica el algoritmo de cluster K-Means a las variables proporción de productos caros adquiridos, porcentaje de productos de precio medio adquiridos y proporción de productos baratos comprados. En este caso aplicamos el algoritmo de cluster a las variables originales sin realizar ningún análisis factorial previo.

Esquema 9.2.2. Metodología de segmentación en función de su sensibilidad al precio .



9.3 RESULTADOS SENSIBILIDAD AL PRECIO A NIVEL GENERAL

La siguiente figura que se adjunta a continuación muestra de forma esquemática un resumen de los segmentos obtenidos. Así observando dicho esquema, se aprecia que se distinguen cuatro segmentos. Dos de ellos se agrupan dentro del grupo de sensibilidad media al precio. Además destaca otro grupo caracterizado por tener una preferencia en la compra por productos baratos por lo que se trata de un segmento con una alta sensibilidad al precio. También aparece el grupo que tiene una fuerte preferencia por comprar productos baratos por lo que se considera que es un grupo insensible al precio a nivel general.

Esquema 9.3.1. Resumen de la segmentación por sensibilidad al precio a nivel general.



La tabla 9.3.2 proporciona más detalles sobre los diferentes segmentos obtenidos. Así se ve como el custer 1 tiene una fuerte preferencia por adquirir productos de gama media. De todos los productos adquiridos por este segmento más del 40 % pertenecen a la categoría de productos de gama media. Este grupo representa el 18,34% de todos los clientes.

Dentro de los segmentos que tienen una sensibilidad media al precio destaca también el segmento 3 ya que es el cluster más numeroso aglutinando a un 44,80% de todos los clientes. Observando el gráfico 9.3.3 parece claro que este segmento es el que más se parece a la distribución general media de los productos vendidos durante el año 2010.

Además según la tabla 9.3.2, el segmento 2 muestra una clara preferencia por productos baratos ya que en promedio la mitad de todos los productos adquiridos por este segmento son baratos.

El segmento 4 destaca por tener una clara preferencia por adquirir productos de un precio alto. En este caso más de la mitad (55,17%) de todos los productos adquiridos son productos considerados como caros.

Tabla 9.3.2. Resumen de la segmentación por sensibilidad al precio a un nivel general.

CLUSTER	% clientes	%_ barato	% medio	% precio alto	Descripción
1	18,34%	29,48%	43,78%	26,74%	Preferencia de productos de gama media.
2	27,02%	50,30%	28,70%	22,00%	Preferencia de productos baratos.
3	44,80%	29,62%	32,97%	37,41%	Compran por igual productos baratos, caros y medios.
4	9,85%	18,75%	26,08%	55,17%	Preferencia de productos caros.

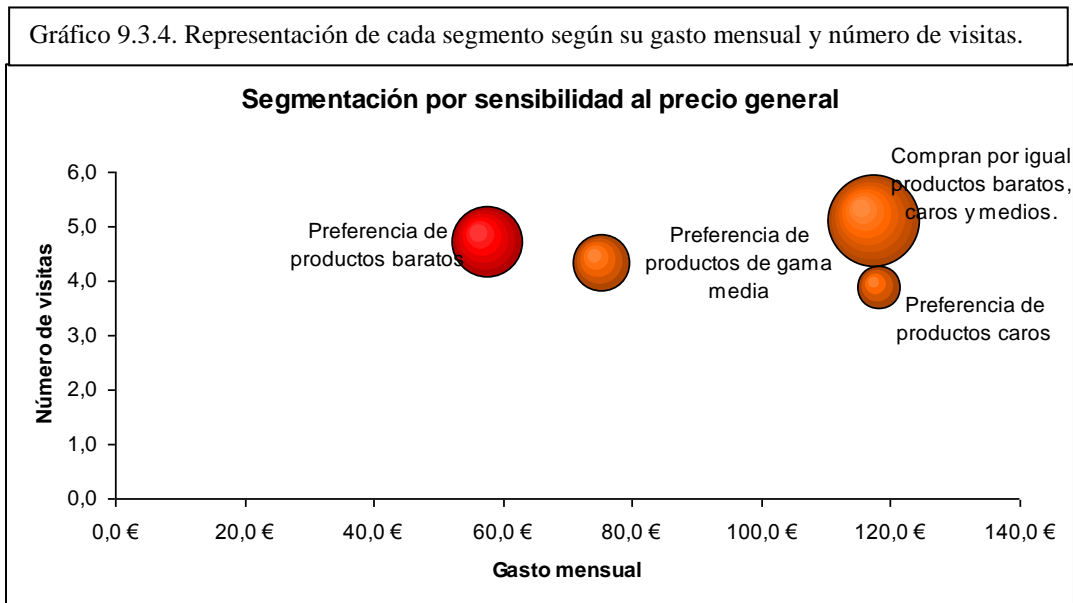
Gráfico 9.3.3. Distribución general de los productos vendidos según la tipología de los productos.

Distribución de los productos vendidos en el año 2010



El gráfico 9.3.3 señala que en términos generales no hay diferencias entre la tipología de los productos adquiridos, es decir, aproximadamente las unidades vendidas de todos los productos caros equivalen a un tercio de todos los productos vendidos.

El gráfico siguiente (gráfico 9.3.4) representa el gasto mensual medio y el número de visitas medio de cada segmento. Además el tamaño de cada segmento es proporcional al porcentaje de clientes que forma parte de cada conglomerado. También el color de cada segmento indica el grado de redención, es decir, el color rojo oscuro señala que es un segmento muy sensible a los cupones mientras que el color naranja indica que es grupo caracterizado por una sensibilidad media al cupón.



Observando el gráfico presentado anteriormente (gráfico 9.3.4) se aprecia que el segmento más numeroso (aquel que compra por igual productos baratos, caros y de gama media) junto con el segmento que tiene una preferencia por productos caros son los dos segmentos que tienen un nivel de gasto más elevado que el resto de los segmentos.

Sin embargo, el segmento que cuenta con una preferencia por productos baratos tiene un gasto medio inferior de los que son menos sensibles al precio a nivel general.

En la figura 9.3.1.2 se presenta un esquema con las características más relevantes de cada segmento.

9.3.1 DISTRIBUCIÓN GENERAL

En este apartado se presenta la distribución global de las variables de segmentación (gasto, visitas, número de compras grandes y redención) y las variables de carácter sociodemográficas para todos los clientes considerados con el objetivo de poder comparar esta distribución general con la distribución de las variables de cada segmento. Con ello se pretende caracterizar a cada segmento.

Gráfico 9.3.1.1 Distribución global.

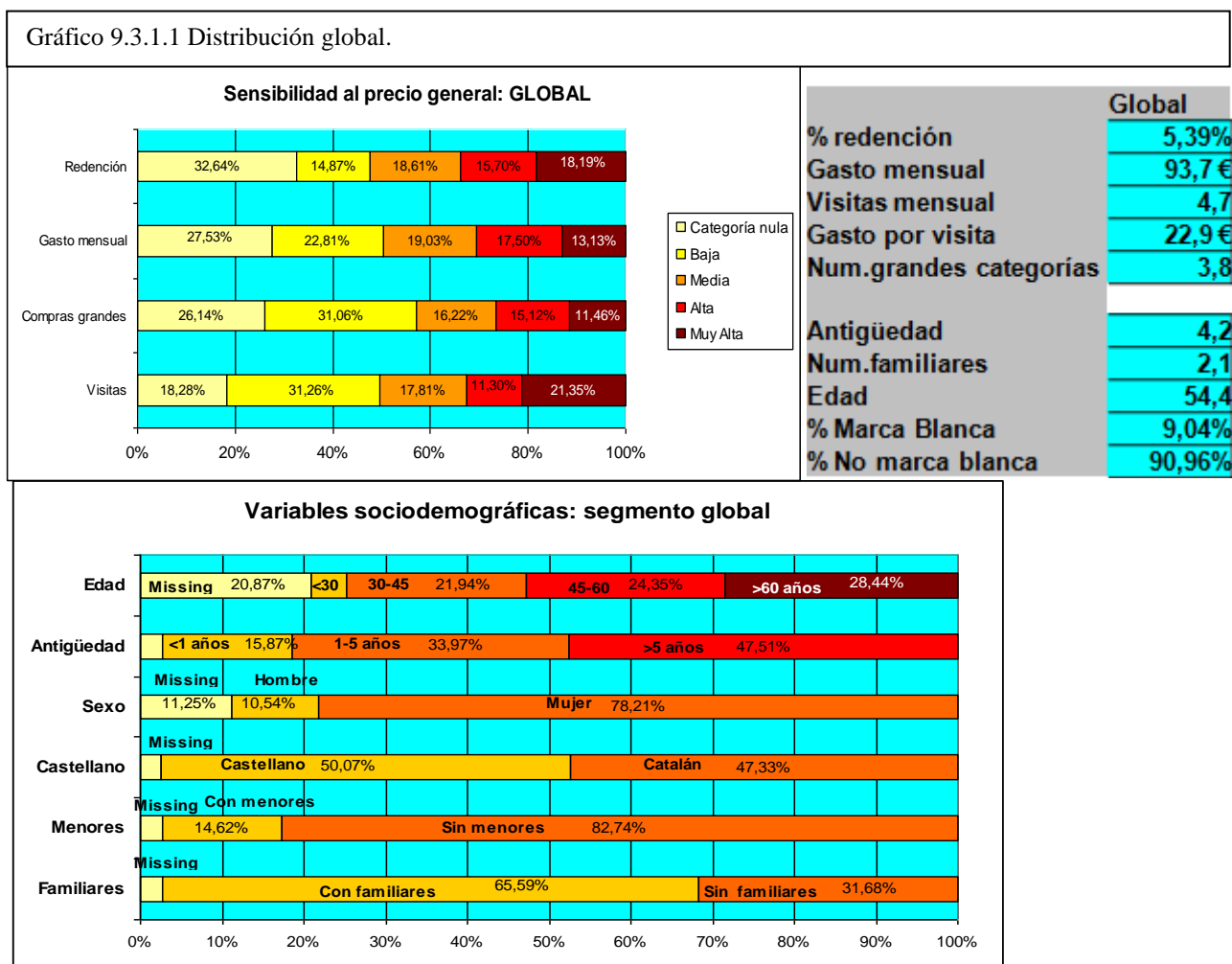
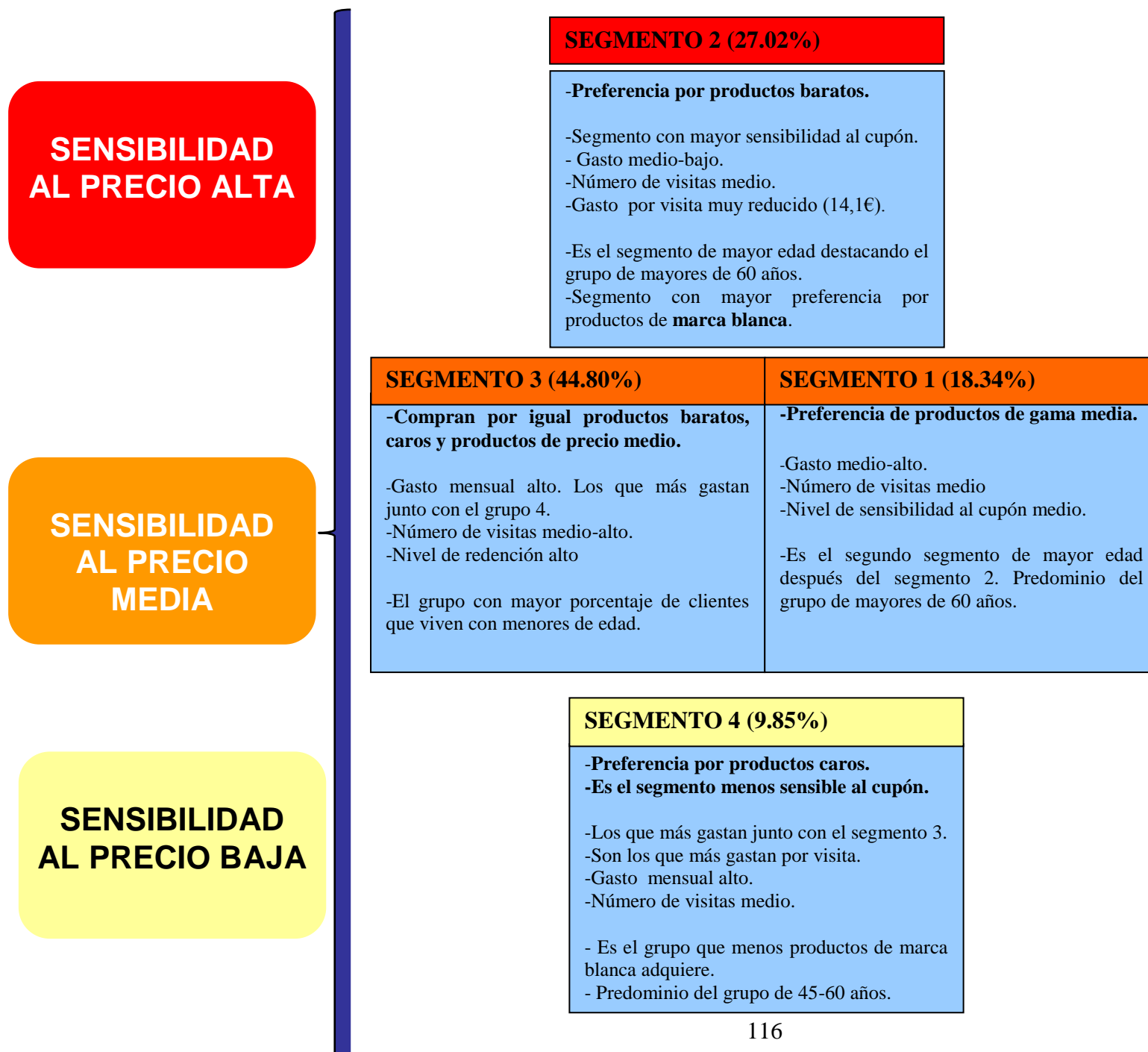


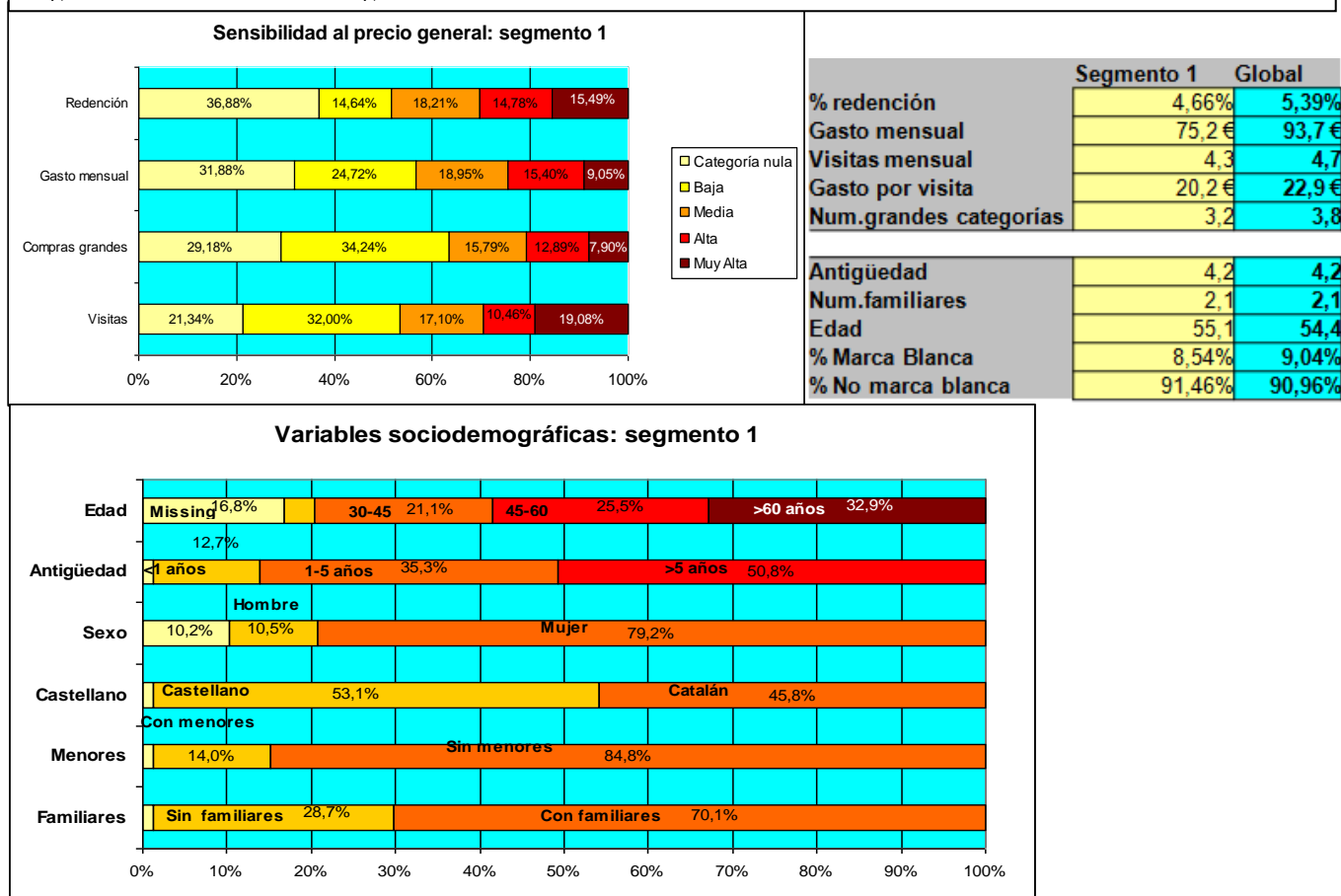
Figura 9.3.1.2. Resumen grupos en función de su nivel de redención.



9.3.2 SEGMENTO 1: "Preferencia de productos de gama media"

Tamaño: 18.34%

Figura 9.3.2. Caracterización segmento 1.



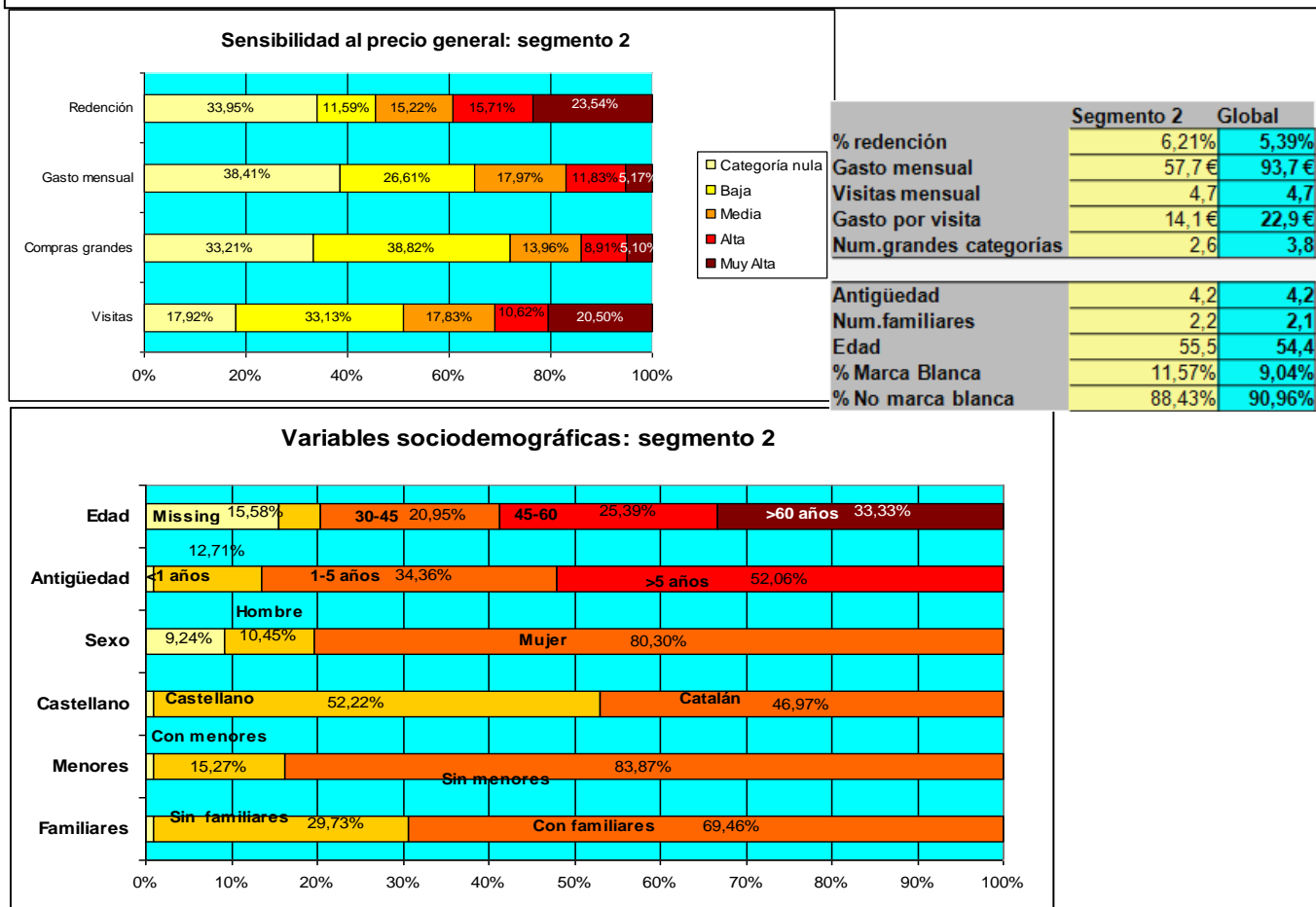
Este grupo, que aglutina al 18,34% de todos los clientes, se caracteriza por lo siguiente:

- Gasto medio-alto, número de visitas medio y nivel de sensibilidad al cupón medio. Además comparado con el segmento 2 tiene un gasto por visita más elevado.
- Es el segundo segmento de mayor edad después del segmento 2. En este segmento destaca el grupo de mayores de 60 años.
- Respecto al porcentaje de marcas blancas adquiridas se asemeja bastante a la media.

9.3.3 SEGMENTO 2: "Preferencia por productos baratos"

Tamaño: 27.02%

Figura 9.3.3. Caracterización segmento 2.



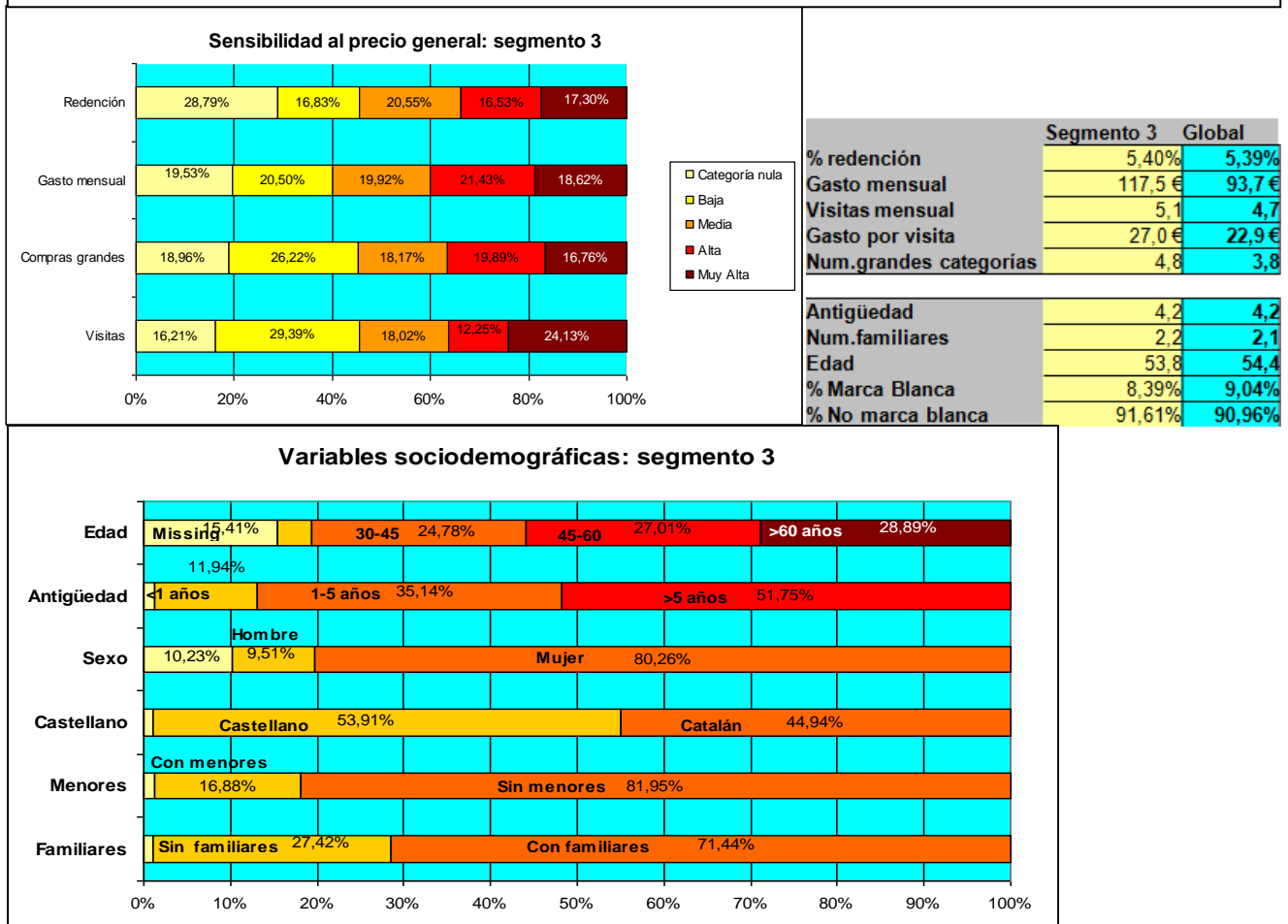
Este grupo, que aglutina al 27,02% de todos los clientes, está definido por las siguientes características:

- Es el segmento con mayor sensibilidad al cupón o mayor nivel de redención.
- Además cuenta con un gasto medio-bajo, número de visitas medio.
- Tiene en promedio un gasto por visita muy reducido (14,1€).
- Es el segmento de mayor edad después del segmento 2 destacando el predominio en grupo de mayores de 60 años.
- Es el segmento con mayor preferencia por productos de marca blanca. De todos los productos adquiridos el 11,57% pertenecen a productos de marca blanca cuando la media supone el 9%.

9.3.4 SEGMENTO 3: "Compran por igual productos baratos, caros y productos de precio medio"

Tamaño: 44.80%

Figura 9.3.4. Caracterización segmento 3.



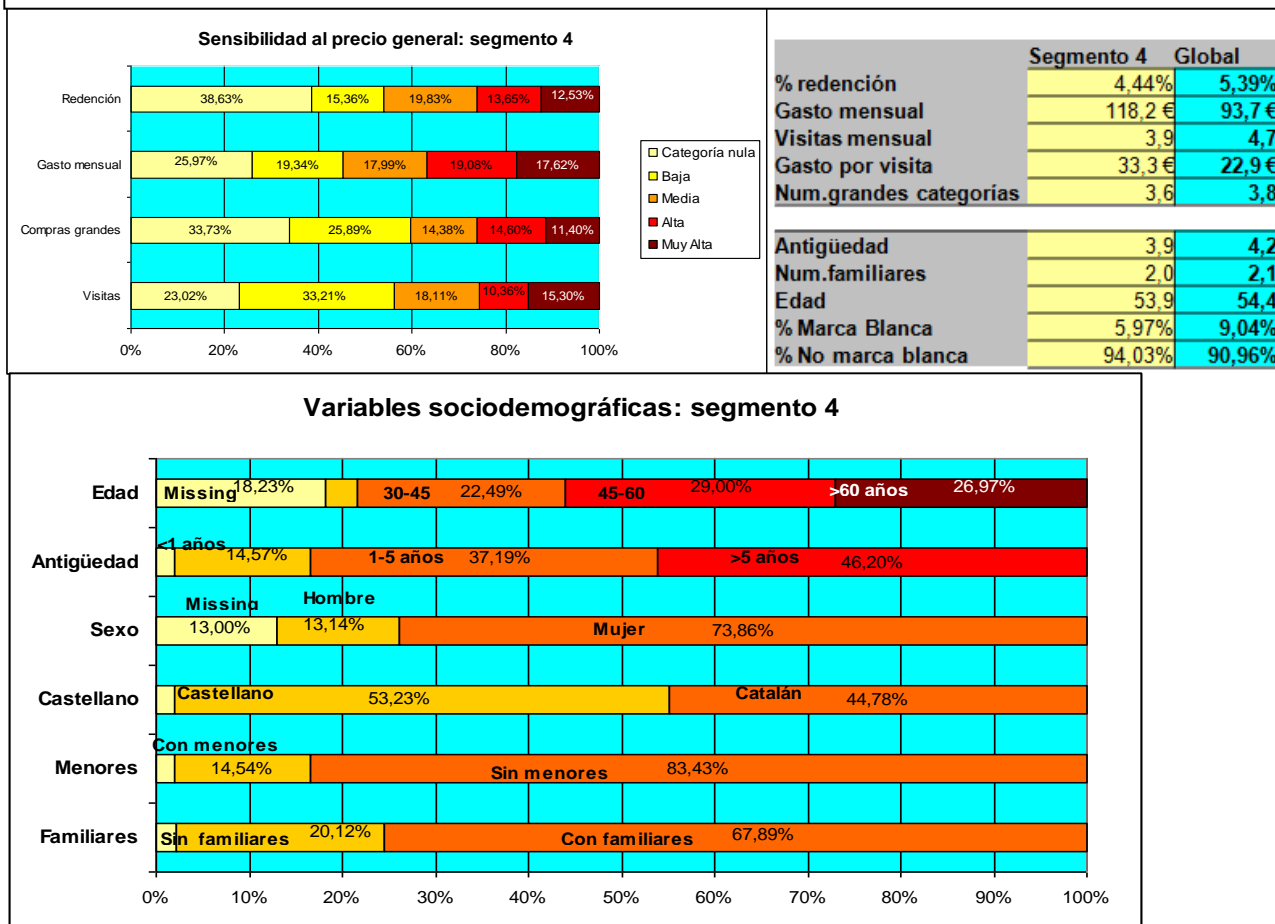
Este grupo, que aglutina al 44,80% de todos los clientes, se caracteriza por lo siguiente:

- Tienen a comprar por igual productos baratos, caros y de precio medio.
- Tienen un gasto mensual alto. Son los que más gastan junto con el grupo 4.
- Cuentan con un número de visitas medio-alto y un nivel de redención alto.
- Tiene un gasto por visita más reducido que el segmento 4.
- Tiene una edad inferior respecto a los primeros dos grupos.
- Destaca por ser el segmento con mayor porcentaje de clientes que viven con menores de edad.

9.3.5 SEGMENTO 4: “Preferencia por productos caros”

Tamaño: 9.85%

Figura 9.3.5. Caracterización segmento 4.



Este segmento aglutina al 9,8% de todos los clientes y se basa en una baja sensibilidad al precio o preferencia por productos de un precio superior. Además se caracteriza por lo siguiente:

- Gasto mensual alto: es el segmento que más gasta junto con el segmento 3. Además es el segmento que más gasta por visita.
- Cuenta con un nivel de redención medio: es el segmento **menos sensible al cupón** ya cuenta con un nivel de redención inferior a otros segmentos.
- Destaca este segmento por ser el grupo que **menos productos de marca blanca adquieren**.
- Respecto a la edad hay que mencionar que existe un predominio del grupo de 45-60 años. Tiene una inferior respecto a los dos primeros grupos.
- Es el segmento que mayor proporción de hombres concentra.

9.4 METODOLOGÍA DE SEGMENTACIÓN EN FUNCIÓN DE LA SENSIBILIDAD AL PRECIO A NIVEL SECCIÓN.

En este apartado se explica de forma breve la metodología empleada en esta segmentación. Esta metodología se puede descomponer en varias etapas y viene esquematizada en el esquema 9.4.3:

1) A partir del número de productos adquiridos por cada cliente en cada sección se calcula la proporción de productos baratos, la proporción de productos de gama media y la proporción de productos caros adquiridos por cada cliente para cada una de las secciones, hasta obtenerse una tabla como la que se muestra a continuación:

Figura 9.4.1.

SECCIÓN 1				SECCIÓN 2		
Id Cliente	Proporción productos precio medio	Proporción productos caros	Proporción productos baratos	Proporción productos precio medio	Proporción productos caros	Proporción productos baratos

- 2) Con la proporción de productos calculada anteriormente para cada sección se **crea una métrica de sensibilidad al precio en una escala porcentual**.
- 3) Tramificamos ese índice de sensibilidad al precio en 4 modalidades tal como se indica en el esquema 9.4.3 o en el apartado de tramificación del índice de sensibilidad 2.6.3.
- 4) Aplicamos un Análisis de Correspondencias Múltiples a los 8 índices de sensibilidad pertenecientes a las 8 secciones. Con los ejes factoriales obtenidas se realiza un cluster jerárquico.

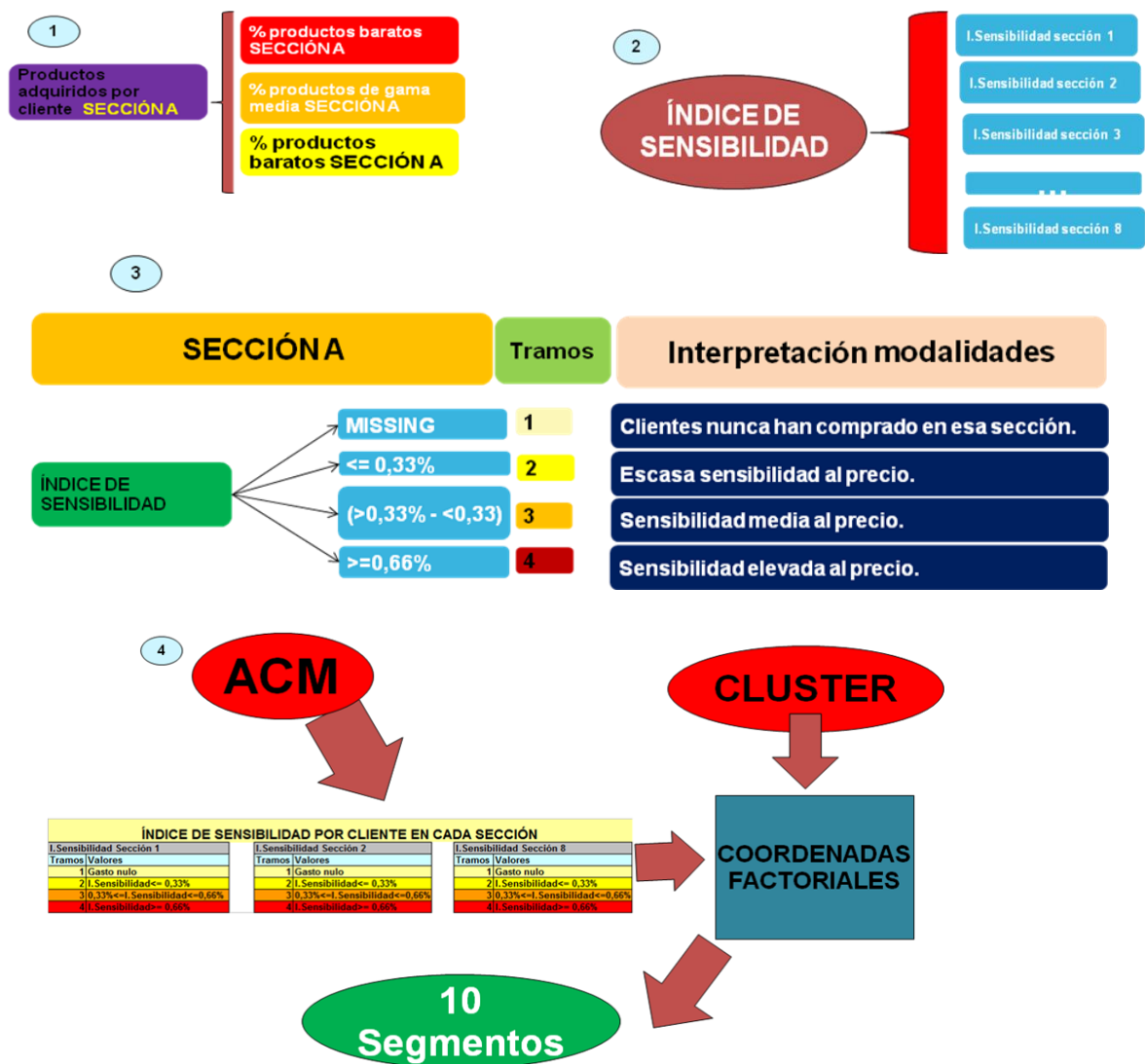
Nota: como el algoritmo de cluster jerárquico no es viable computacionalmente cuando se tiene un volumen elevado de individuos, se ha seleccionado una muestra que representa el 10% de toda la población mediante un muestreo aleatorio simple.

Además solamente se tiene en cuenta para el cómputo de la sensibilidad al precio las compras efectuadas en las siguientes secciones que muestra la tabla 9.4.2.

Tabla 9.4.2.

Secciones consideradas	
Alimentación seca	Perfumería
Líquidos	Panadería
Cremería	Lácteos
Conservas	Congelado

Esquema 9.4.3. Resumen de la Metodología empleada.



9.5 RESULTADOS

9.5.1 ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

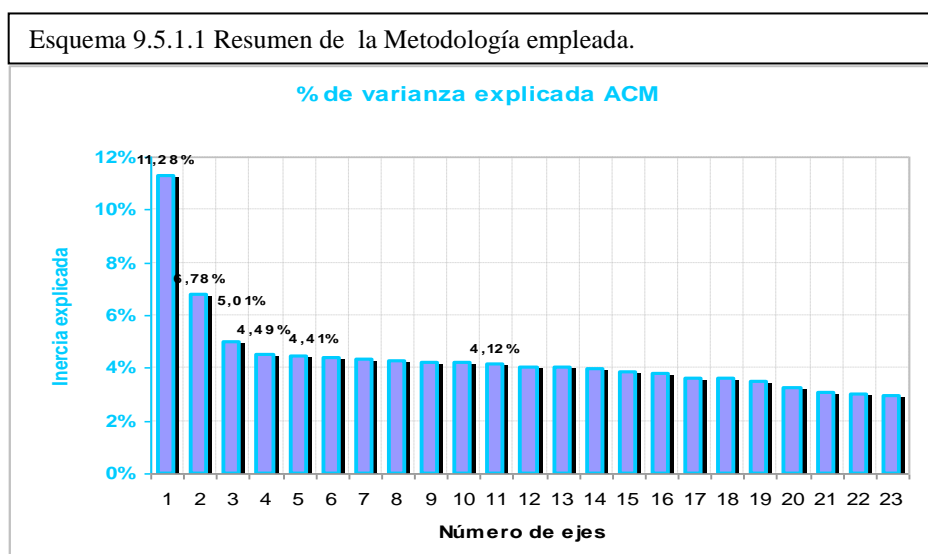
El Análisis de Correspondencias Múltiple que se presenta en este apartado ha sido realizado tanto en SAS como en SPAD. En este apartado se presentará el ACM realizado en SPAD.

Para interpretar el Análisis de Correspondencias Múltiple seguimos el mismo esquema empleado en el apartado anterior:

- 1) Análisis de la descomposición de la inercia explicada por cada eje factorial.
- 2) Interpretación gráfica de los ejes factoriales:
 - a. Proyección de las variables activas sobre los ejes factoriales.
 - b. Proyección de las variables suplementarias sobre los ejes factoriales.
 - c. Proyección de la nube de puntos de los individuos junto con las variables activas

1) TABLA DE DESCOMPOSICIÓN DE LA VARIANZA.

Con el siguiente gráfico representamos la descomposición de la varianza explicada por cada eje. En él se puede observar que el primer eje factorial recoge gran parte de la inercia siendo con gran diferencia el eje más diferenciado.



En el histograma se aprecia que los dos primeros ejes recogen aproximadamente el 18% de toda la inercia, sobresaliendo entre ellos el primer eje con una explicabilidad del 11,28%.

2) INTERPRETACIÓN GRÁFICA DE LOS EJES FACTORIALES.

En el gráfico 9.5.1.2 se proyectan las 8 variables activas referentes al índice de sensibilidad por cliente para las 8 secciones consideradas en esta segmentación.

Se aprecia claramente en el gráfico 9.5.1.2 que en la parte de la derecha se concentran aquellas modalidades de gasto nulo en cada sección, mientras que en la parte de la izquierda aparecen las modalidades que indican un nivel de sensibilidad al precio medio. En la parte superior (zona 2) destacan las modalidades que hacen referencia a una sensibilidad reducida al precio, es decir, en esa zona del plano factorial estarán aquellos clientes que suelen comprar productos caros. En la parte inferior del gráfico se sitúan las categorías correspondientes a una alta sensibilidad al precio, es decir, en esa zona del plano aparecerán aquellos clientes que tienden a comprar productos baratos.

Con esta información **se puede interpretar el eje 2** como un eje que hace referencia a la **preferencia por productos caros**. De esta forma, el eje 2 opone clientes que tienden a comprar productos caros contra aquellos que compran productos baratos.

Para interpretar mejor el eje 1 es necesaria la ayuda del gráfico 9.5.1.3 en el que se proyecta en el plano factorial las variables suplementarias de fidelidad (gasto, número de visitas, número de compras grandes y redención). Analizando el gráfico se desprende que en la parte de la derecha destacan las modalidades de fidelidad reducida en cada sección (gasto inferior a 20€, menos de una visita al mes, redención nula y 0 compras grandes) mientras que en la izquierda aparecen aquellas modalidades relacionadas con un gasto alto, un elevado número de visitas y un número elevado de compras grandes. Por lo tanto a la vista de esto, **se puede interpretar el eje 1** como un **eje de fidelidad**. Además se repite lo comentado en las segmentaciones anteriores: las variables gasto mensual, número de visitas y número de compras grandes siguen la misma tendencia pero la variable redención se escapa de ese efecto. Con esto parece que la tendencia es la siguiente: clientes con un nivel de sensibilidad al precio medio tienden a tener un nivel de fidelidad alto mientras que aquellos clientes que tienen cierta preferencia por adquirir productos baratos tienen un nivel de fidelidad más reducido.

Volviendo a la figura 9.5.1.2 se aprecia que las trayectorias de las 8 variables activas siguen el mismo esquema aunque se distingue que la sección congelado se aparta un poco de esa tendencia. La modalidad de gasto nulo de sección congelado se sitúa más hacia la izquierda (fidelidad más alta) que las demás secciones. Esto quiere decir que esta sección es menos competitiva que las demás secciones en el sentido de que habrá más clientes que no compren en congelado que en otras secciones. Todo lo contrario ocurre para la sección alimentación seca en la que su modalidad de gasto nulo se encuentra más apartada siendo la más alejada.

Resulta curioso contemplar el gráfico 9.5.1.4 en el que se compara la trayectoria de la variable sensibilidad al precio en la sección congelado con la variable suplementaria nivel redención. En este gráfico se aprecia que las dos variables siguen un esquema muy similar. Además parece que existe asociación entre ellas, es decir, clientes con un nivel de redención alto tienden a ser muy sensibles al precio en la sección congelado.

También parece que, para todas las secciones, niveles reducidos de redención se asocian con niveles bajos de sensibilidad al precio mientras que niveles altos de redención están más asociados con niveles altos de sensibilidad al precio. Con esto parece que aquellos

clientes que son menos sensibles a los cupones tienen una preferencia más alta a comprar productos caros que aquellos clientes con una sensibilidad más alta que tienden a comprar productos más baratos.

En las figuras 9.5.1.5 y 9.5.1.6 se muestra en detalle ampliando el zoom la proyección de las variables socio-demográficas. En términos generales estas variables no son muy significativas a la hora de describir los ejes ya que todas las modalidades se concentran en el centro de la nube de puntos. Aunque no hay grandes diferencias entre ellas parece que los jóvenes (menores de 30 años) son más sensibles al precio que los otros grupos de mayor edad. Además se aprecia lo comentado en otras segmentaciones: modalidades referentes a una antigüedad inferior a un año, vivir en el hogar sin menores de edad, vivir en el hogar sin familiares y ser catalanoparlante están asociadas a un nivel de fidelidad más reducido.

FIGURA 9.5.1.2: Proyección de las variables activas y su trayectoria.

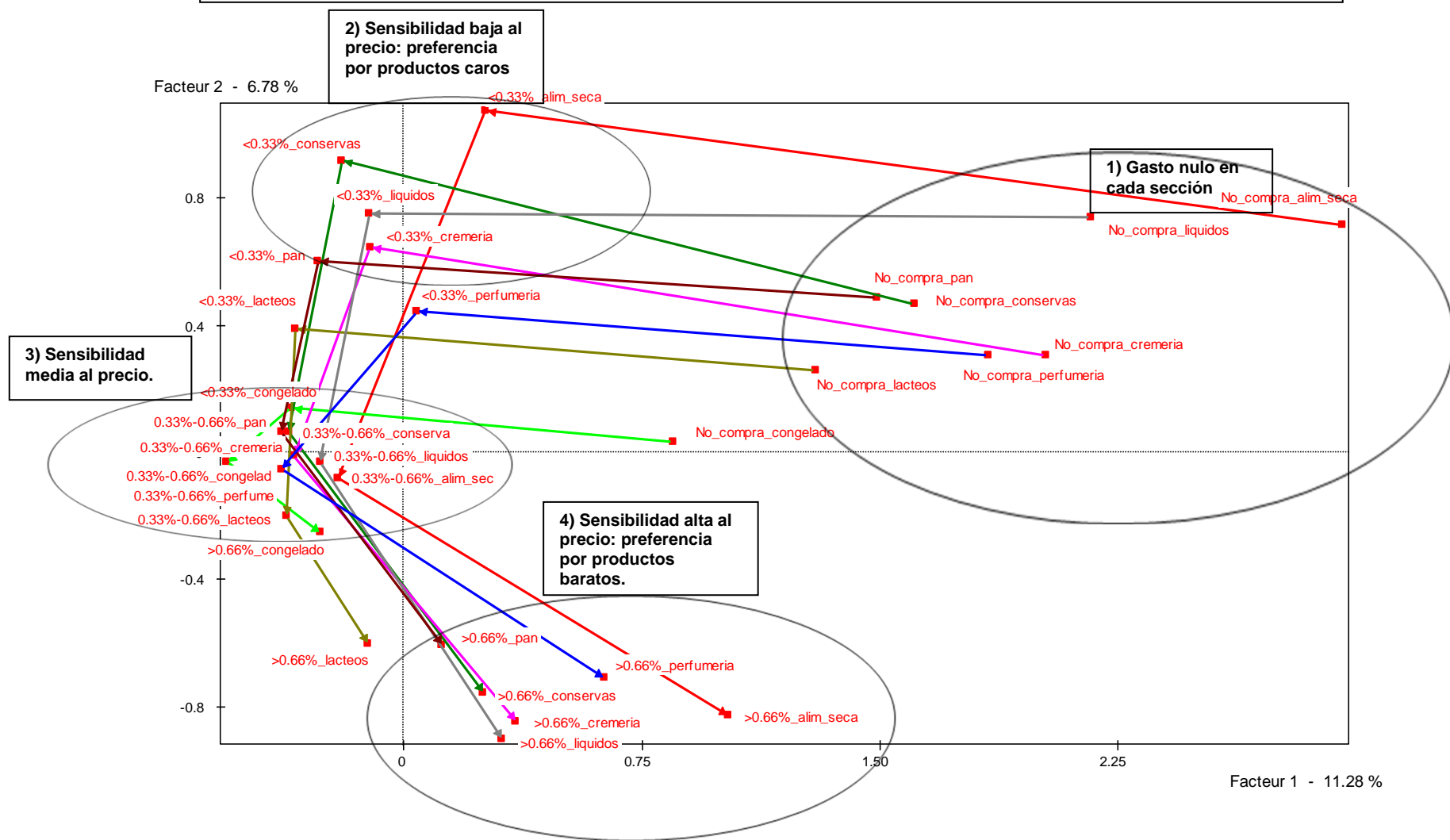


FIGURA 9.5.1.3: Proyección de las variables suplementarias de fidelidad (gasto, visitas, número de compras grandes y redención)

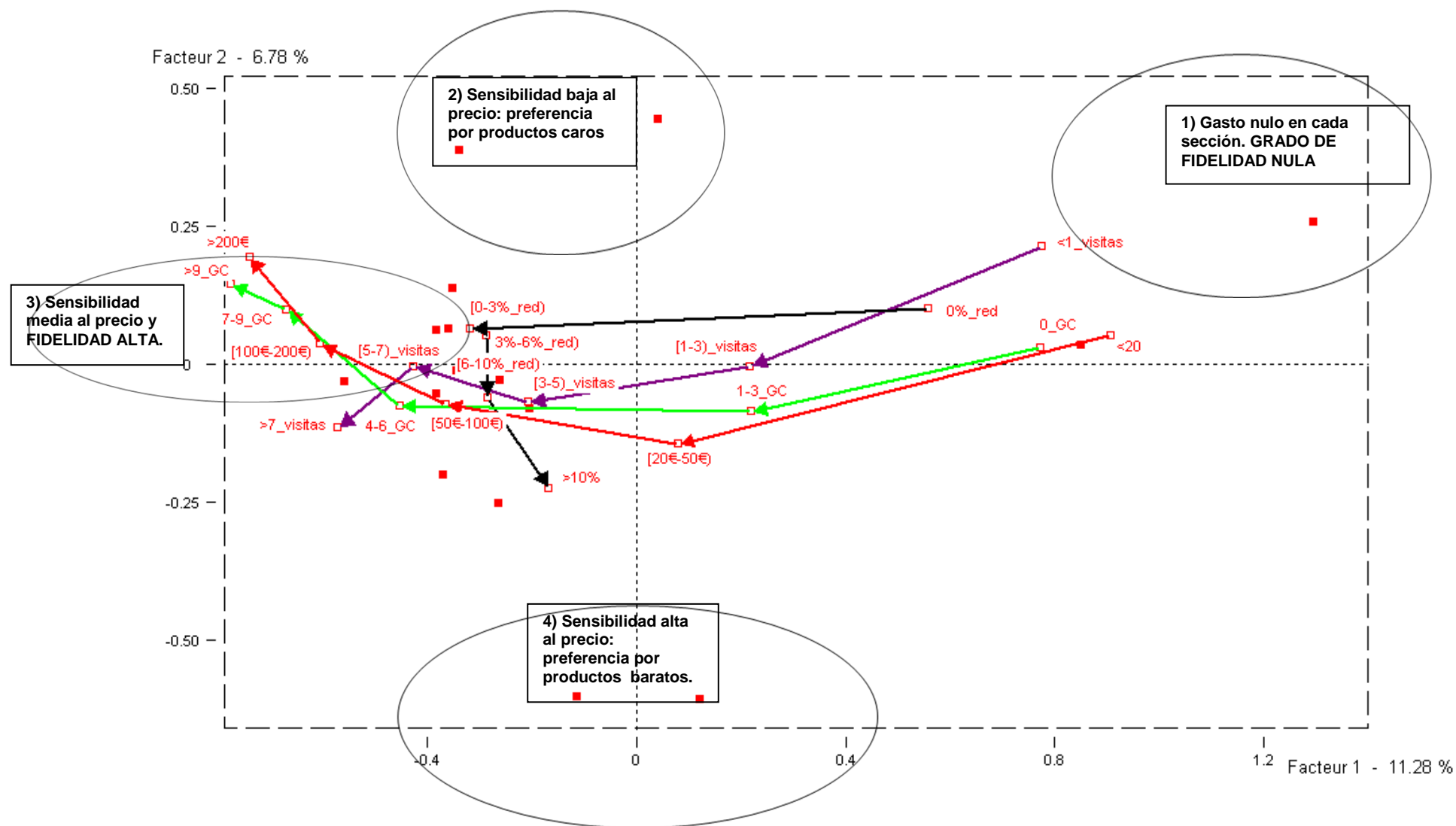


FIGURA 9.5.1.4: Comparación de la trayectoria de la variable activa índice de sensibilidad de la sección congelado con la variable suplementaria nivel de redención.

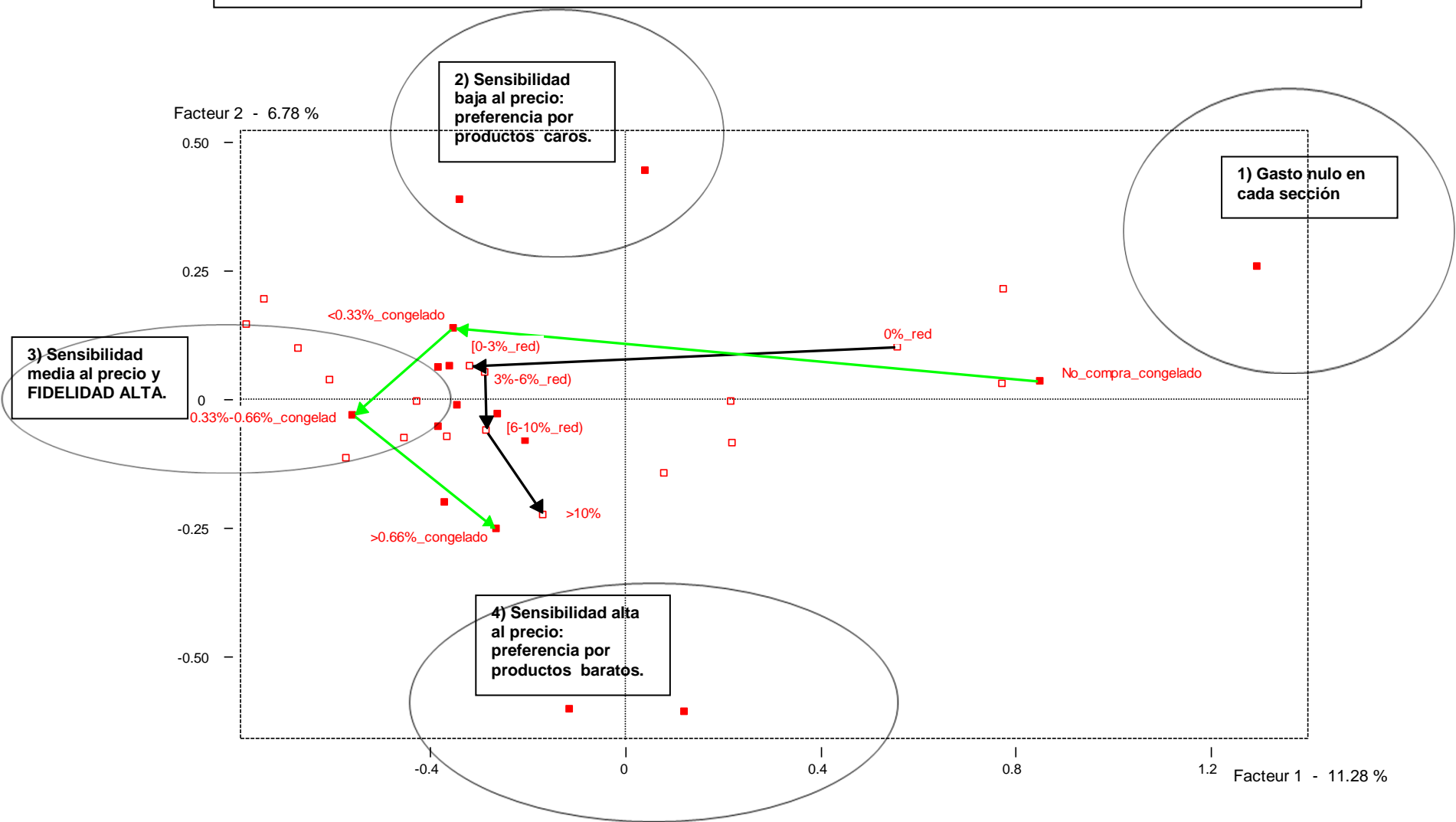


FIGURA 9.5.1.5: Proyección de las variables suplementarias sociodemográficas: edad, si vive con menores, idioma y antigüedad.

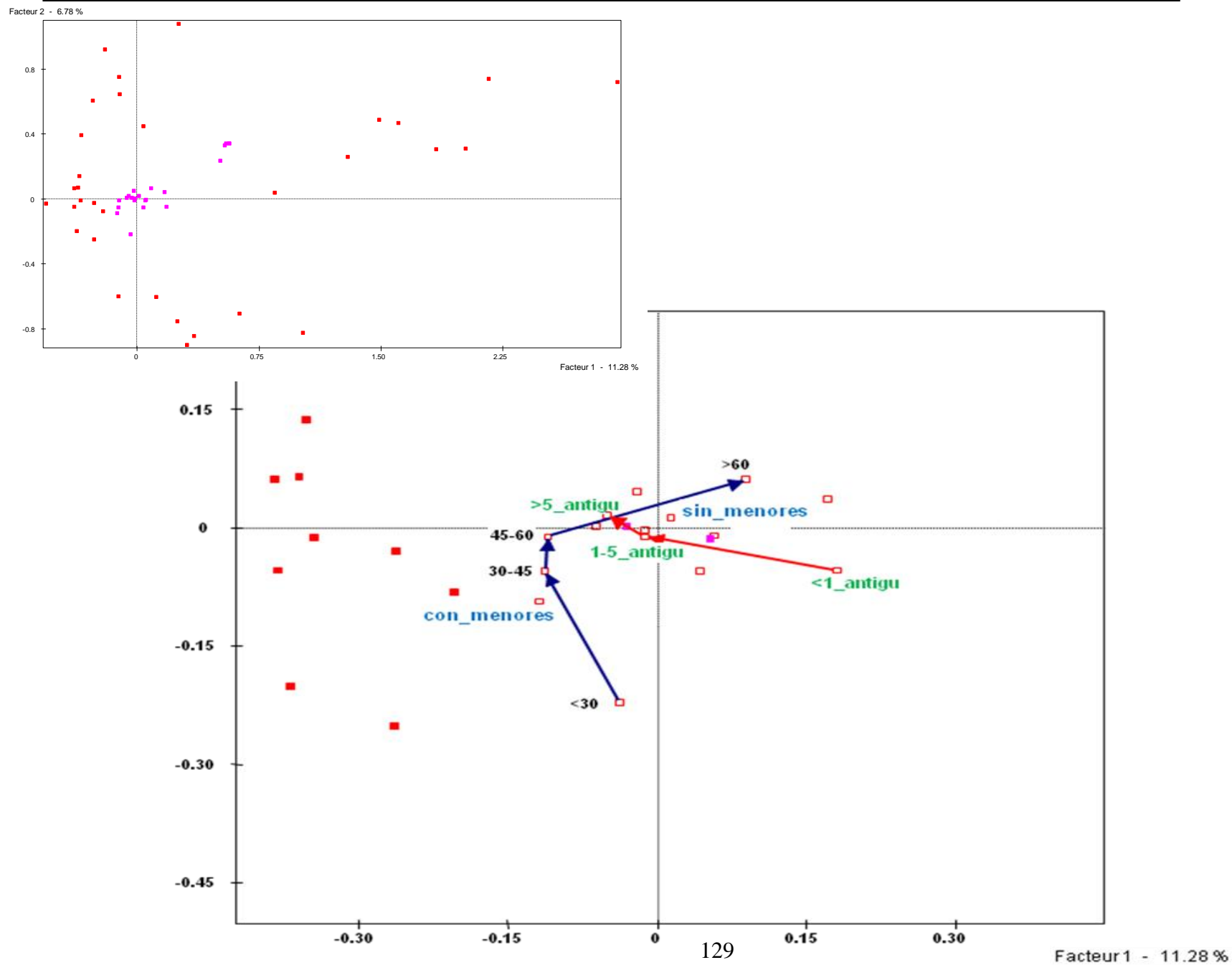


FIGURA 9.5.1.6: Proyección de las variables suplementarias sociodemográficas: edad, sexo, si vive con menores, idioma y antigüedad.

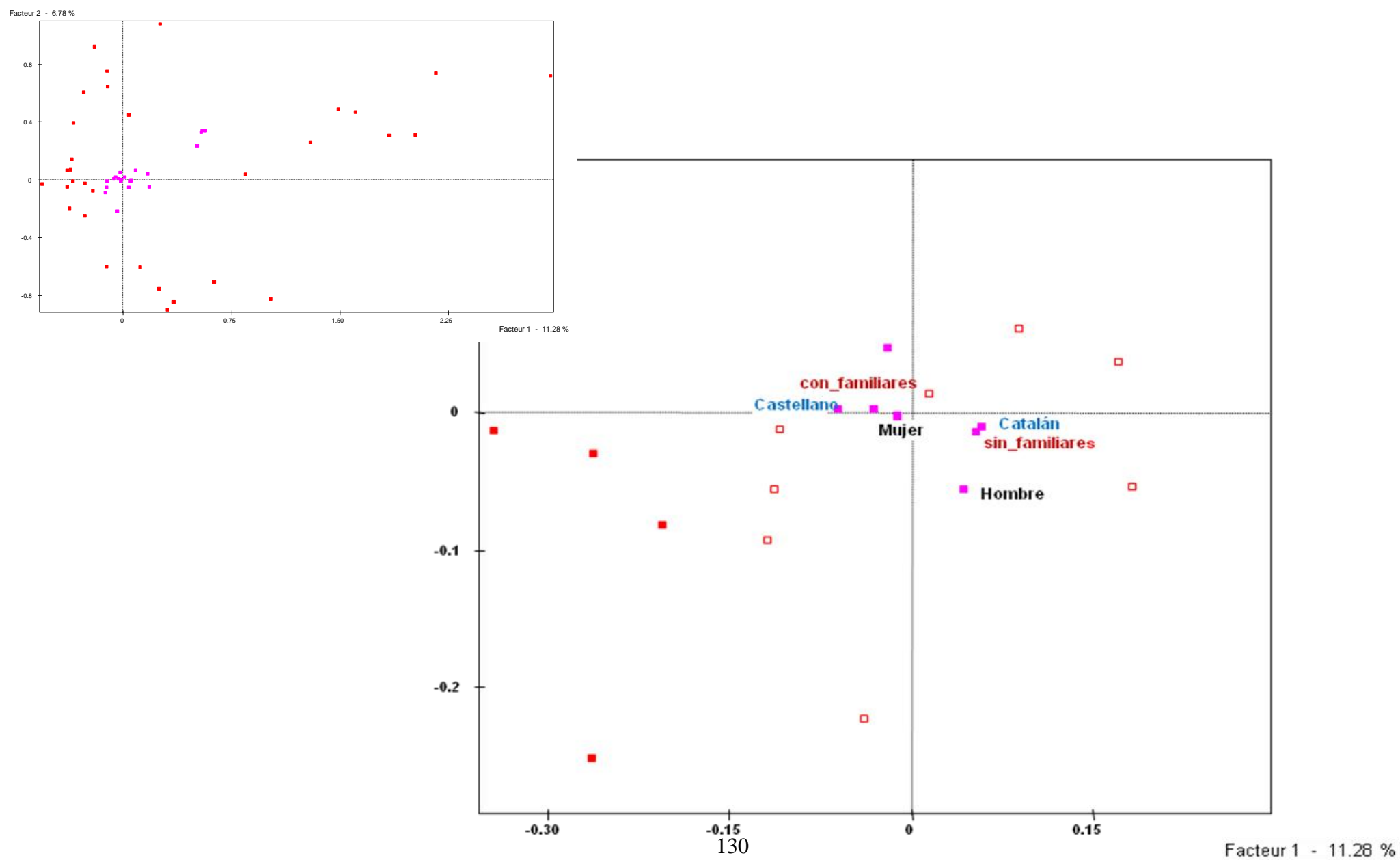
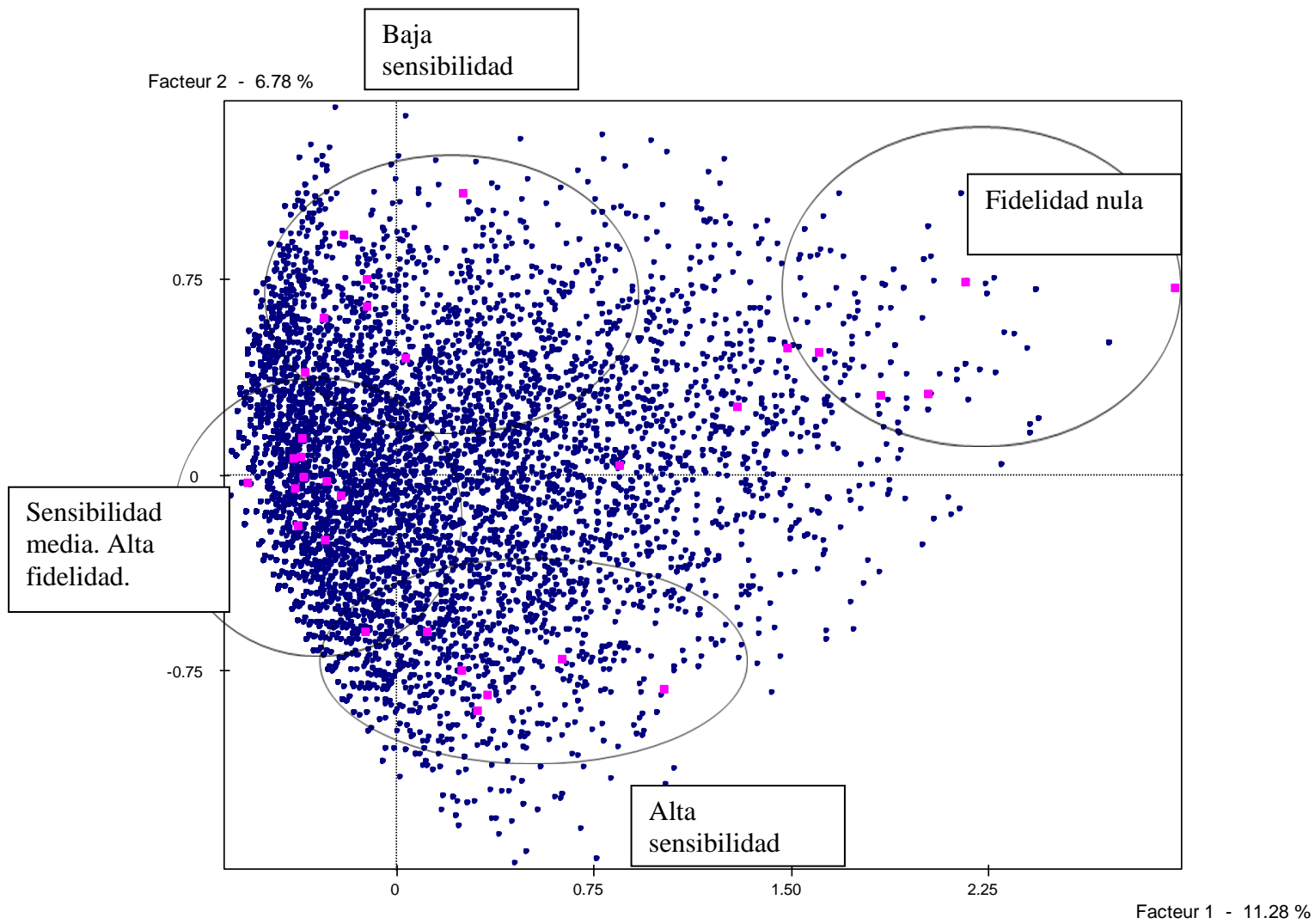


FIGURA 9.5.1.7: Proyección de las variables suplementarias sociodemográficas: edad, sexo, si vive con menores, idioma y antigüedad.

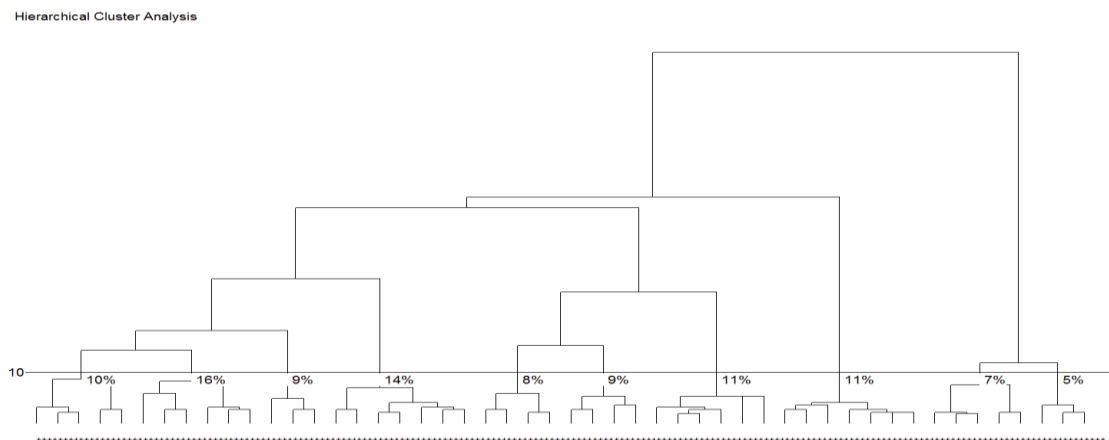


9.5.2 RESULTADOS CLUSTER JERÁRQUICO

Después de realizar el Análisis de Correspondencias Múltiples, en este apartado se presentan los resultados de la segmentación tras aplicar el algoritmo de cluster jerárquico a los 10 primeros ejes factoriales. Esta segmentación ha sido realizada con el software multivariante de estadística SPAD.

En primer lugar se muestra el dendograma obtenido a partir de la aplicación del cluster jerárquico con el criterio de Ward. Como se puede ver en el gráfico 9.5.2.1 se han considerado 10 grupos.

Esquema 9.5.2.1. Dendograma obtenido a partir del método jerárquico



La figura 9.5.2.2 que se adjunta proporciona una visión general y esquemática de todos los segmentos obtenidos. En el esquema se ordenan los segmentos en función de su grado de sensibilidad al precio: de mayor a menor sensibilidad al precio.

Es importante aclarar en el esquema el significado de los colores: el color rojo significa un nivel de sensibilidad al precio elevado, el naranja un nivel medio de sensibilidad al precio, el amarillo un nivel bajo de sensibilidad al precio y el gris un nivel muy reducido de gasto o gasto nulo.

Como se observa en el esquema se pueden agrupar a los segmentos en cuatro categorías o subgrupos:

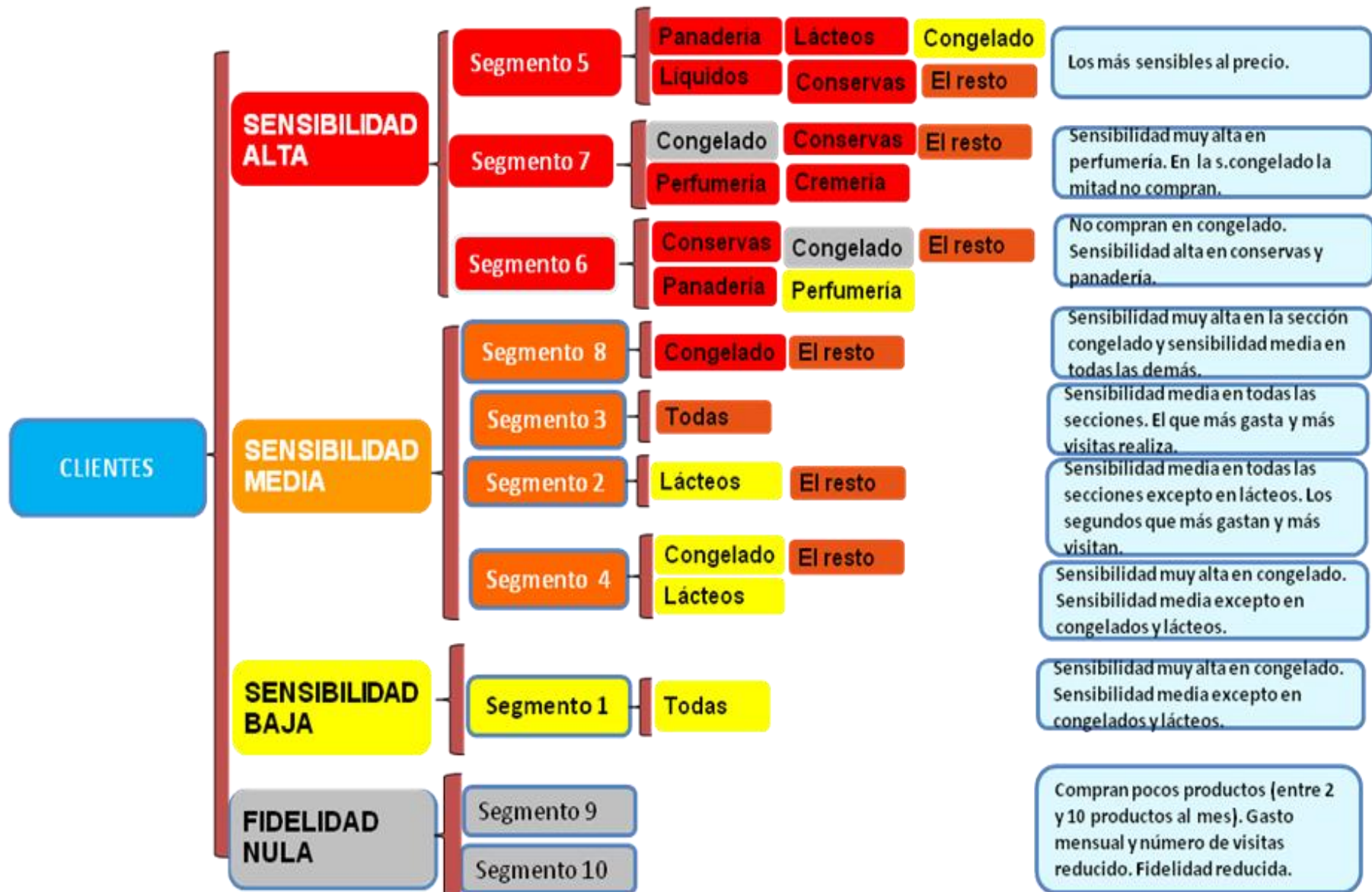
- 1) Los que son muy sensibles al precio y tienen preferencia por productos baratos.
- 2) Los que tienen una sensibilidad media al precio.
- 3) Los más insensibles al precio (color amarillo).
- 4) Los que tienen un grado de fidelidad nula (color gris).

Observando la proyección de los clusters obtenidos sobre el espacio factorial junto con las variables activas y las variables suplementarias de gasto mensual y nivel de redención (Gráfico 9.5.2.3 y Gráfico 9.5.2.4) se aprecia claramente que aquellos **clusters caracterizados por tener un nivel medio de sensibilidad al precio en casi todas las secciones tienen un nivel de fidelidad más elevado**, es decir, son grupos que tienen un gasto mensual y un número de visitas más elevado que otros segmentos. Esto se aprecia en el gráfico 9.5.2.5 en el que los grupos con mayor nivel de gasto y con mayor número de visitas son aquellos caracterizados por tener un nivel medio de sensibilidad al precio.

Respecto a los grupos caracterizados por tener un nivel de sensibilidad al precio elevado destacan tres segmentos:

- El segmento 5, que es el más sensible al precio, se caracteriza por comprar productos baratos en las secciones de panadería, lácteos, líquidos, conservas, por ser insensible en la sección congelado y por tener un nivel medio de sensibilidad en las secciones restantes.
- El segmento 7, destaca por tener una sensibilidad muy alta en la sección perfumería y porque la mitad de los clientes del segmento no compran en la sección congelado. Además tienen una sensibilidad relativamente alta en las secciones conservas y cremería y sensibilidad media en el resto.
- El segmento 6 se caracteriza por no comprar en la sección congelado, por una sensibilidad alta en conservas y panadería y además tienden a comprar productos caros en la sección perfumería.

Esquema 9.5.2.2. Esquema sensibilidad al precio a nivel sección.



Para los grupos caracterizados por un nivel medio de sensibilidad al precio existen cuatro conglomerados:

- Segmento 8: tiene un nivel de sensibilidad muy alto en la sección congelado y un nivel de sensibilidad medio en el resto de secciones.
- Segmento 3: tiene un nivel de sensibilidad medio en todas las secciones.
- Segmento 2: caracterizado por una sensibilidad media en todas las secciones excepto en la sección lácteos donde optan por productos más caros.
- Segmento 4, que compra productos baratos en la sección congelados y lácteos y tiene un nivel de sensibilidad medio en las demás secciones.

Además de estos grupos existen dos segmentos (segmento 9 y 10) que tienen en común su bajo nivel de fidelidad.

Gráfico 9.5.2.3. Proyección de los clusters sobre el espacio factorial junto con las variables activas.

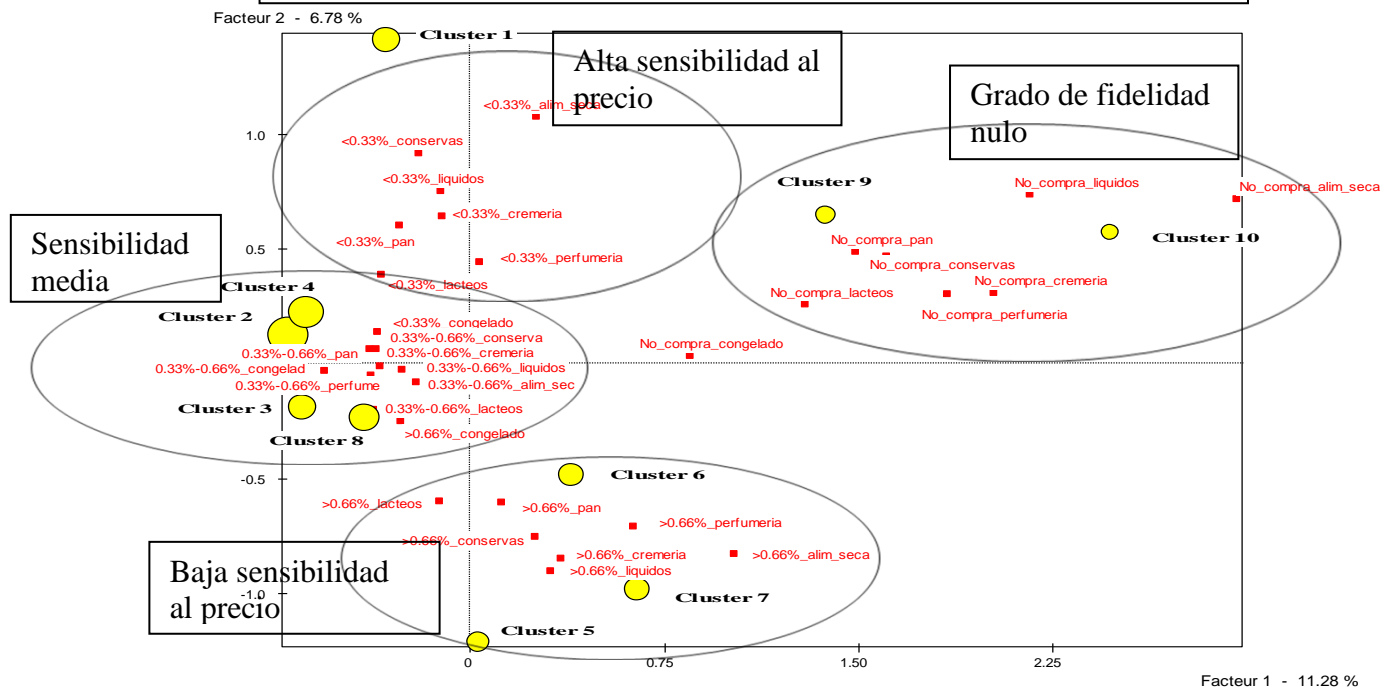


Gráfico 9.5.2.4. Proyección de los clusters sobre el espacio factorial junto con las variables suplementarias gasto mensual y nivel de redención.

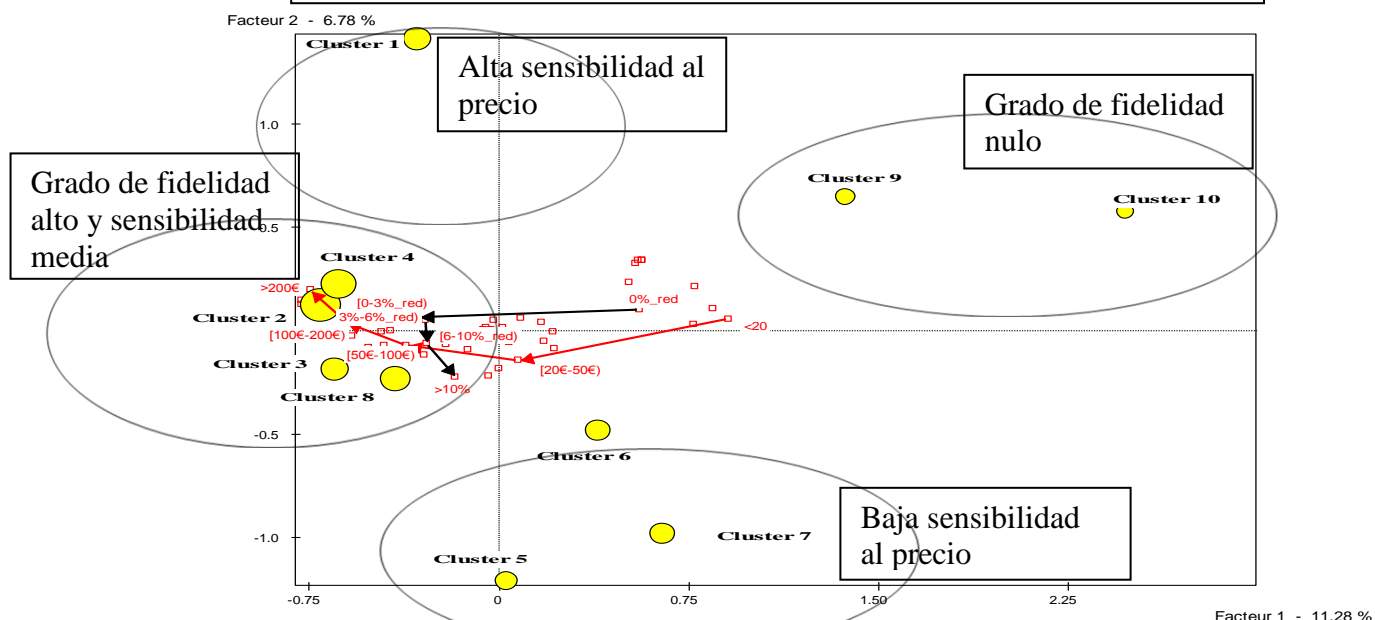
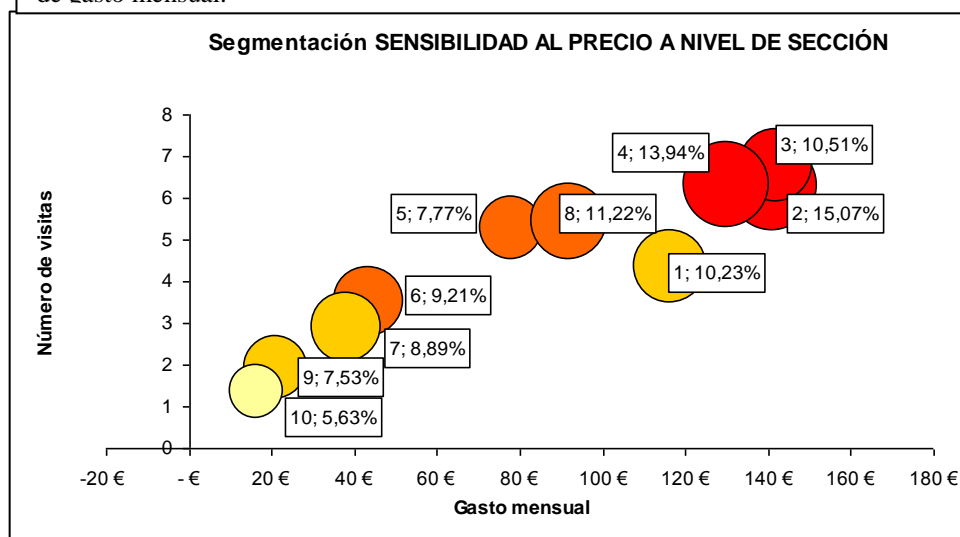
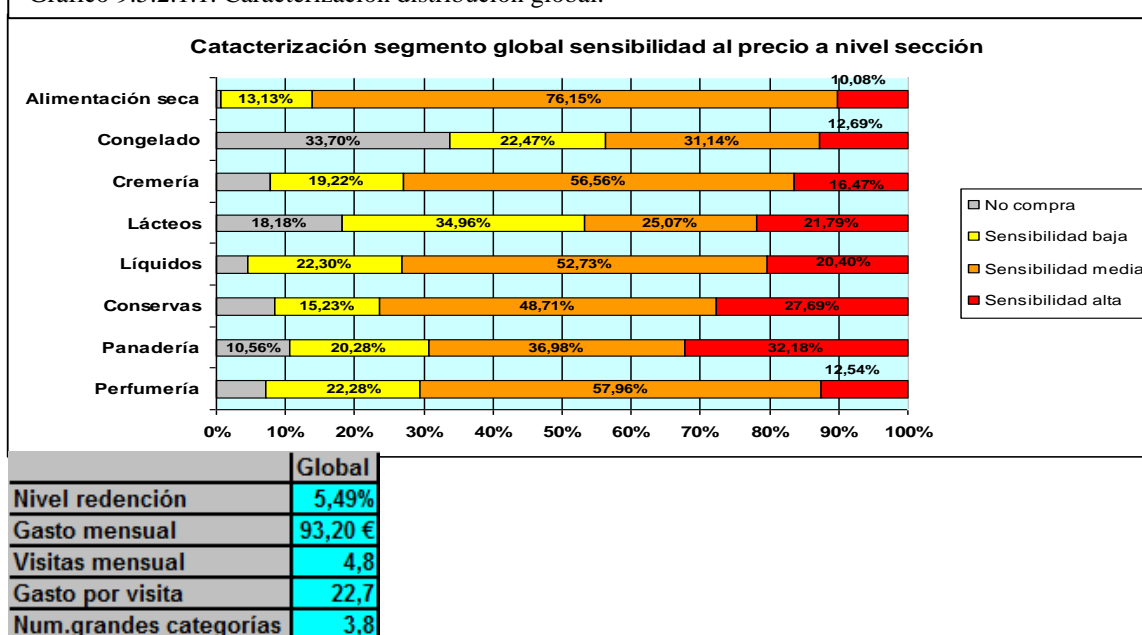


Gráfico 9.5.2.5. Representación de los 10 cluster en función de su nivel de visitas y de gasto mensual.



9.5.2.1 DISTRIBUCIÓN GENERAL

Gráfico 9.5.2.1.1. Caracterización distribución global.



En este apartado mostramos el nivel de sensibilidad general para cada sección. Como se aprecia en el gráfico anterior la mayor parte de los clientes (76,15%) tienen un nivel medio de sensibilidad al precio para la sección alimentación seca. Además se observa que los clientes son más sensibles al precio en la sección panadería y conservas y más insensibles en la sección de lácteos. También se aprecia en el gráfico que un 33% de todos los clientes no realizan ningún gasto en la sección congelado.

A continuación se presentará cada segmento en detalle empezando por los más sensibles al precio.

Figura 9.5.2.2. Resumen de segmentos obtenidos con la segmentación sensibilidad al precio a nivel de sección.

**SENSIBILIDAD
AL PRECIO ALTA**

SEGMENTO 5 (7,77%)

- Es el segmento más sensible al precio.
- Tiende a comprar productos baratos en: panadería, lácteos, líquidos y conservas.
- Tiene un nivel de sensibilidad bajo al precio en la sección congelado y un nivel medio en las demás secciones.
- Nivel de sensibilidad al cupón alto (superior al 6%).
- Gasto mensual medio
- Número de visitas alto.
- Gasto por visita es reducido.

SEGMENTO 7 (8,89%)

- Todos los clientes de este segmento compran productos baratos en la sección perfumería.
- Además tiene sensibilidad alta en conservas y cremería y sensibilidad media en el resto de secciones.
- Nivel de sensibilidad a los cupones medio.
- Gasto bajo (20€-50€).
- Número de visitas medio.

SEGMENTO 6 (9,21%)

- Nivel elevado de sensibilidad en: Cremería, Panadería y Conservas
- No compran en congelado y tienden a comprar productos caros en perfumería.
- En el resto de secciones tiene un nivel de sensibilidad medio.
- Gasto bajo (20€-50€) y un número de visitas medio
- Gasto por visita es reducido.

SEGMENTO 1 (10.23%)

- Es el segmento con un nivel de sensibilidad al precio más reducido: tendencia a adquirir productos caros, especialmente en la sección cremería.
- Tiene un nivel de gasto mensual elevado y un nivel de visitas medio.
- Es el segmento que más gasta por visita.
- Sensibilidad al cupón media.

**SENSIBILIDAD
AL PRECIO BAJA**

SEGMENTO 9 (7.53%)

- Es el segundo segmento que menos gasta.
- Tienen un gasto y un número de visitas muy reducido, siendo el segundo grupo que menos gasto realiza.

SEGMENTO 10 (5.63%)

- Es el segmento con menor grado de fidelidad: el que menos gasta al mes y el que menos visita.
- El índice de sensibilidad para estos clientes no es muy significativo ya que han comprado pocos productos en secciones: entre 2 y 10 productos al mes.

FIDELIDAD NULA

Figura 9.5.2.2. Resumen de segmentos obtenidos con la segmentación sensibilidad al precio a nivel de sección.

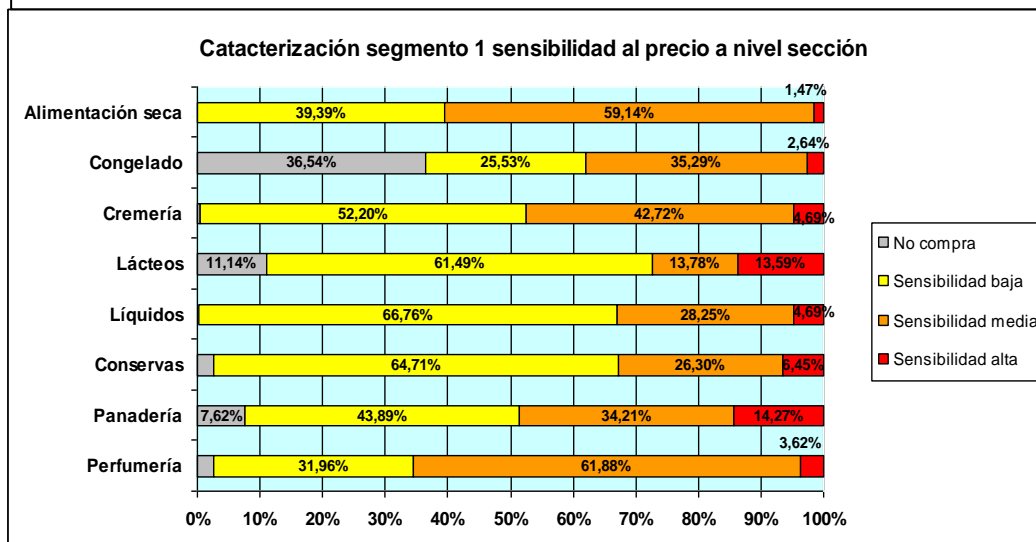
**SENSIBILIDAD
MEDIA AL
PRECIO**

SEGMENTO 8 (11.22%) - Sensibilidad muy alta en congelado y media en las demás secciones -Gasto mensual y un número de visitas medio. -Nivel de sensibilidad medio a los cupones.	SEGMENTO 3 (10.51%) -Nivel de sensibilidad medio en todas las secciones. Es el segmento que más gasto mensual y número de visitas realiza. - Nivel de redención alta. - Segmento caracterizado por vivir con menores de edad y con familiares en el hogar.
SEGMENTO 2 (15.07%) -Sensibilidad medio en todas las secciones excepto en la sección lácteos donde por término medio existe una preferencia a adquirir productos caros. -Es el segundo cluster que más gasta y el que más visitas realiza. -Tiene un nivel de redención alta.	SEGMENTO 4 (13.94%) -Sensibilidad media en todas las secciones excepto en la sección congelado y en la sección lácteos. -Tiene un gasto mensual elevado (más de 100€ mensuales) y un número de visitas elevado. -Nivel de redención medio.

9.5.2.2 SEGMENTO 1 “Los menos sensibles al precio: tendencia a adquirir productos caros”

Tamaño: 10.23%

Figura 9.5.2.2 Caracterización segmento 1.



	Segmento 1 Global	
Nivel redención	4,72%	5,49%
Gasto mensual	116,4 €	93,2 €
Visitas mensual	4,3	4,8
Gasto por visita	31,7 €	22,7 €
Num-grandes categorías	3,7	3,8

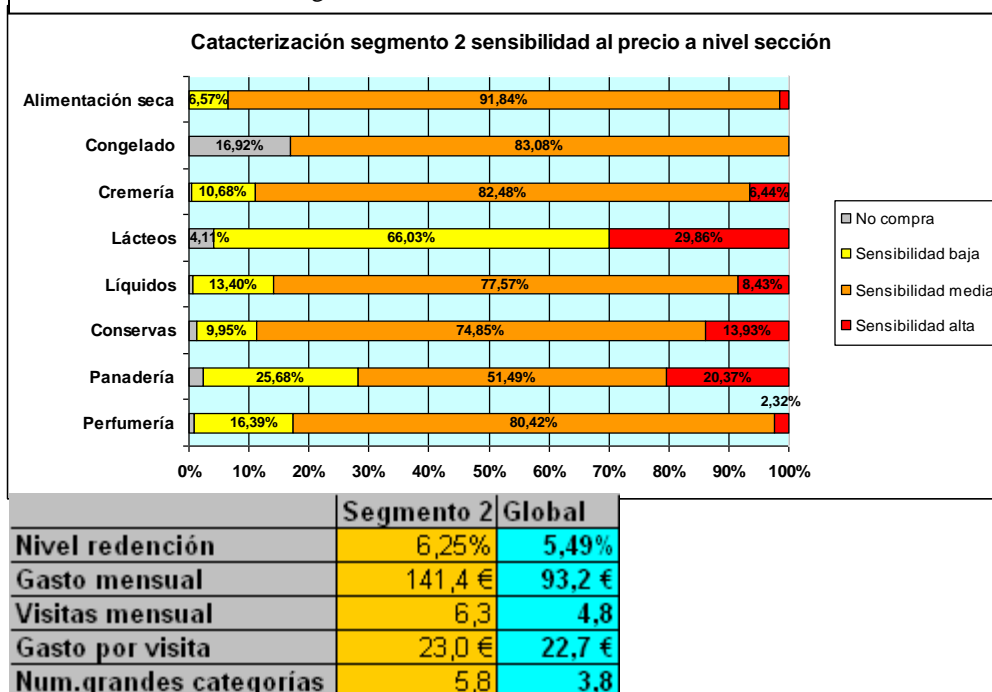
Este segmento representa el 10,23% de todos los clientes. Viene descrito por las siguientes características:

- Es el segmento con un nivel de sensibilidad al precio más bajo. Para casi todas las secciones se ve que hay una tendencia a adquirir productos caros, especialmente en la sección cremería.
- En general se puede decir que su nivel de sensibilidad al precio es el más reducido de todos los segmentos.
- Tiene un nivel de gasto mensual elevado y un nivel de visitas medio.
- Es el segmento que más gasta por visita.
- Tiene un nivel medio de sensibilidad al cupón.

9.5.2.3 SEGMENTO 2 “Sensibilidad media en todas las secciones excepto en lácteos”

Tamaño:15.07 %

9.5.2.3. Caracterización segmento 2



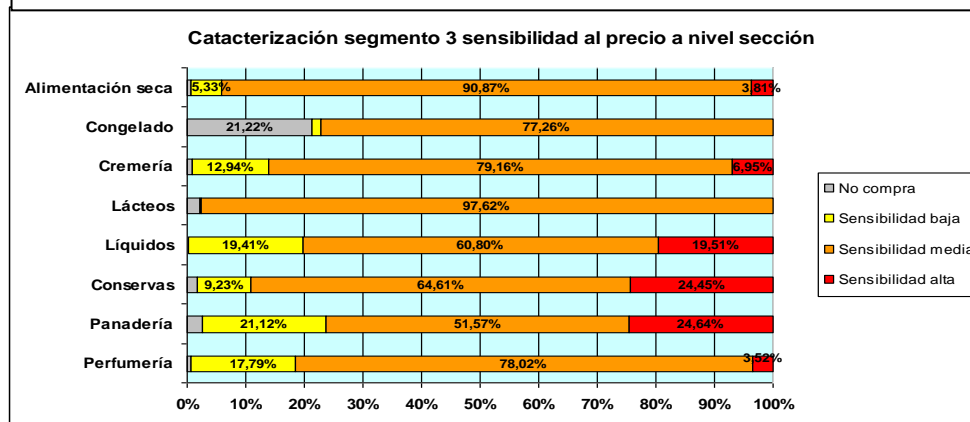
El segmento 2 tiene un tamaño del 15,07% de todos los clientes y se identifica por las siguientes características:

- Se caracteriza por tener un nivel de sensibilidad medio en todas las secciones excepto en la sección lácteos donde por término medio existe una preferencia a adquirir productos caros.
- En la sección lácteos la mayoría (el 66%) tiende a comprar productos baratos pero existe casi un 30% de los clientes que tienen un nivel de sensibilidad alto.
- Es el segundo cluster que más gasta y que más visitas realiza. Tiene un nivel de gasto y un nivel de visitas elevado.
- Además cuenta con un nivel de redención alto.

9.5.2.4 SEGMENTO 3” Sensibilidad media en todas las secciones”

Tamaño: 10.51 %

9.5.2.4. Caracterización segmento 3.



	Segmento 3	Global
Nivel redención	6,37%	5,49%
Gasto mensual	141,9 €	93,2 €
Visitas mensual	6,8	4,8
Gasto por visita	25,0 €	22,7 €
Num.grandes categorías	5,7	3,8

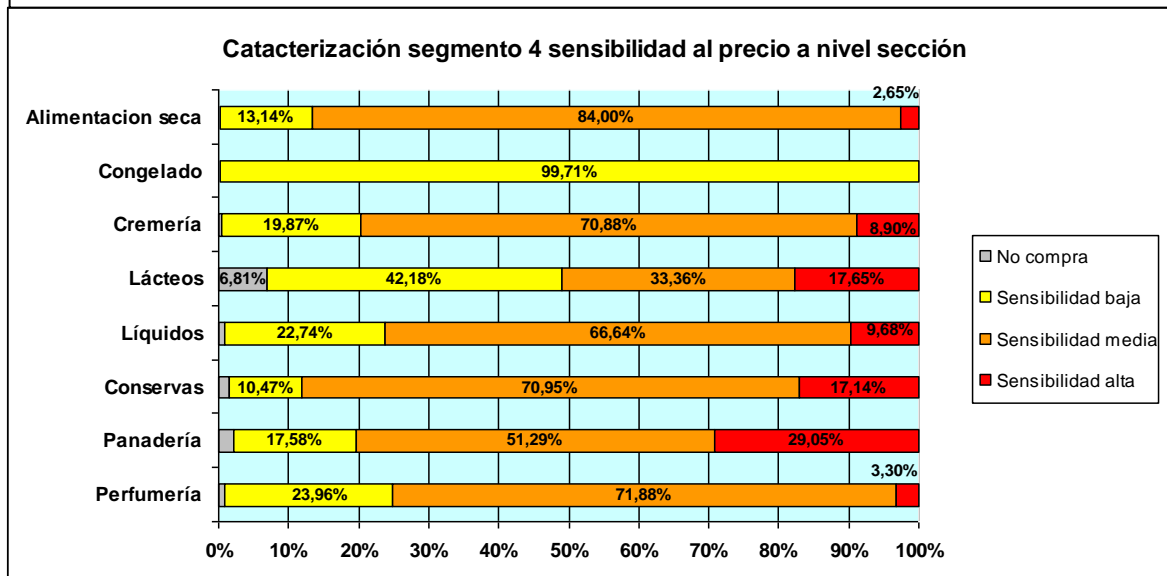
El segmento 3, con un tamaño del 10,51% de todos los clientes, se identifica por:

- Un nivel de sensibilidad medio en todas las secciones.
- Es el segmento que más gasto mensual y número de visitas realiza.
- Tiene un nivel de redención alto.
- Es un segmento caracterizado por vivir con menores de edad y con familiares en el hogar.

9.5.2.5 SEGMENTO 4 “Sensibilidad media excepto en congelado y lácteos”

Tamaño: 13.94 %

9.5.2.5. Caracterización segmento 4

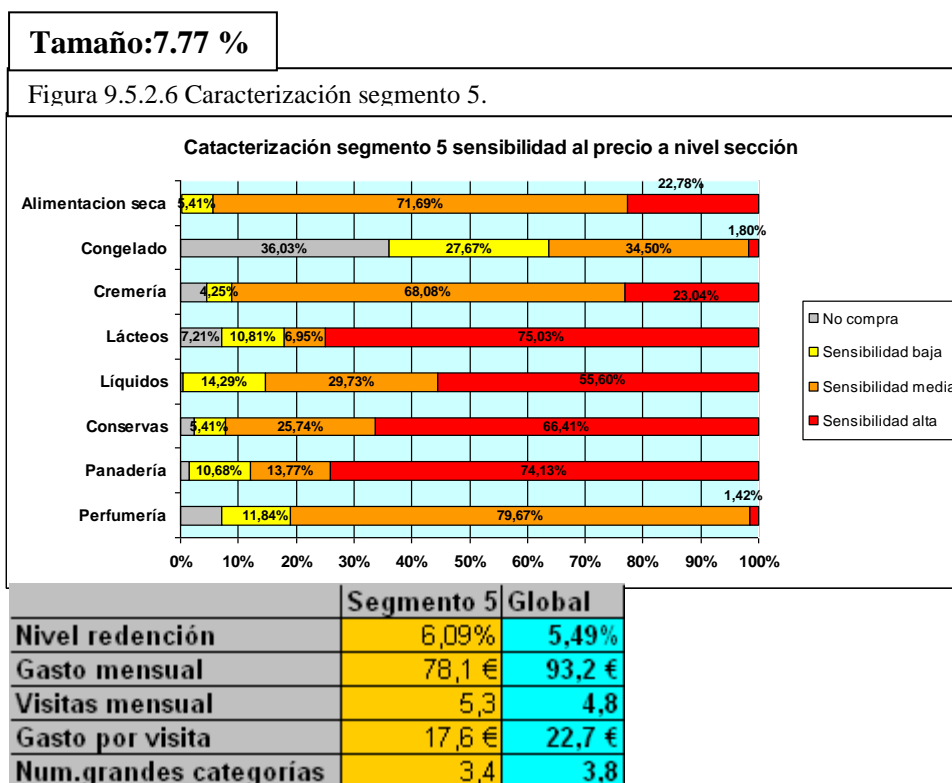


	Segmento 4	Global
Nivel redención	6,05%	5,49%
Gasto mensual	130,0 €	93,2 €
Visitas mensual	6,3	4,8
Gasto por visita	25,9 €	22,7 €
Num.grandes categorías	5,5	3,8

El segmento 4 tiene un tamaño del 13,94% de todos los clientes y se caracteriza por:

- Tiene un nivel de sensibilidad medio en todas las secciones excepto en la sección congelado y en la sección lácteos.
- Todos los clientes que pertenecen a este cluster tienden a comprar productos caros en la sección congelado.
- En la sección lácteos tienen un nivel de sensibilidad medio-bajo lo que supone una ligera preferencia por adquirir productos caros en la sección de lácteos.
- Tiene un gasto mensual elevado (más de 100€ mensuales) y un número de visitas elevado.
- Respecto a la sensibilidad al cupón hay que decir que es un segmento con una sensibilidad media.

9.5.2.6 SEGMENTO 5 “Los más sensibles al precio”



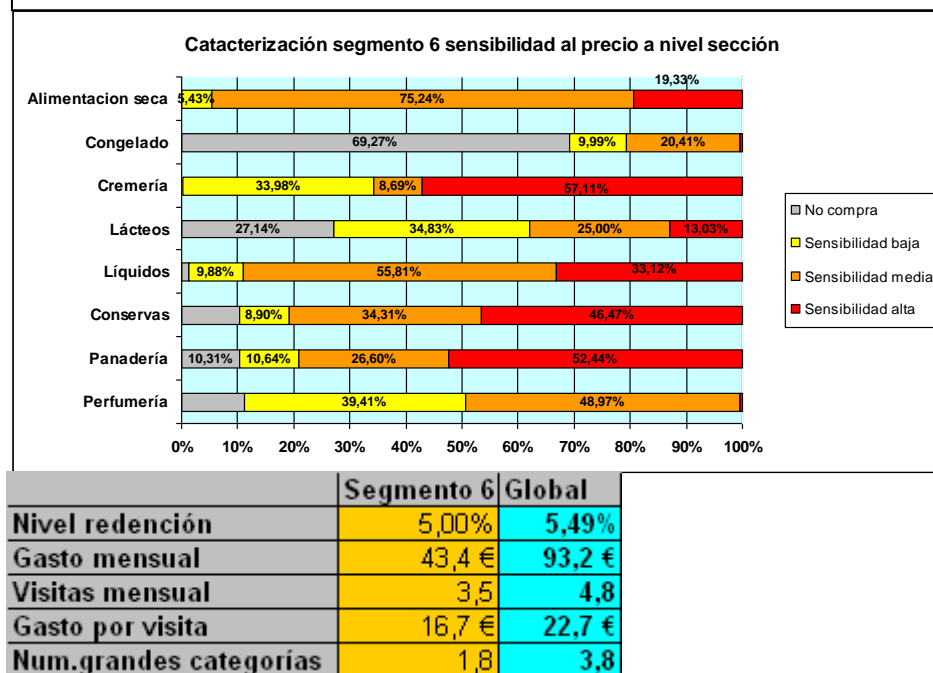
El segmento 5, que representa el 7,77% de todos los clientes, se caracteriza por lo siguiente:

- Es el segmento más sensible al precio.
- Tiende a comprar productos baratos en las siguientes secciones:
 - Panadería.
 - Lácteos.
 - Líquidos
 - Conservas.
- Tiene un nivel de sensibilidad bajo al precio en la sección congelado.
- En las demás secciones tiene un nivel medio de sensibilidad al precio.
- Cuenta con un nivel de sensibilidad al cupón alto (superior al 6%).
- Tiene un gasto mensual medio y un número de visitas elevado, es decir, es un segmento caracterizado por realizar muchas visitas respecto a su nivel de gasto mensual. Por lo tanto su gasto por visita es reducido.

9.5.2.7 SEGMENTO 6 “No compra en la sección congelados y sensibilidad baja en perfumería”

Tamaño: 9.21%

Figura 9.5.2.7 Caracterización segmento 6.



El segmento 6, compuesto por el 9,21% de todos los clientes, se caracteriza por lo siguiente:

-El 70% no compra en la sección congelado.

-Existe sensibilidad alta en:

- Sección Cremería.
- Sección Panadería.
- Sección Conservas.

-Aunque es un grupo con una sensibilidad al precio considerable en varias secciones, en la sección perfumería optan por consumir productos más caros.

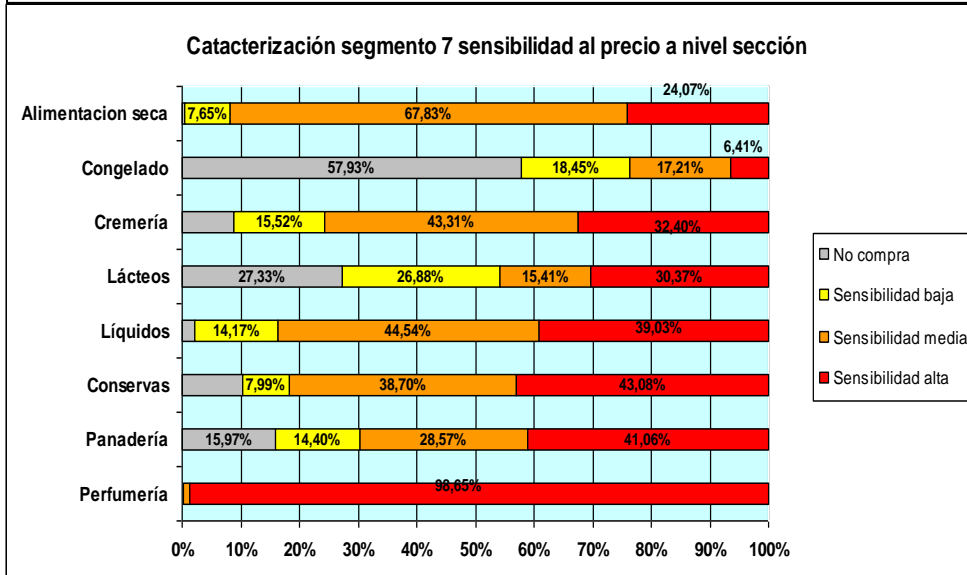
-En el resto de secciones tiene un nivel de sensibilidad medio.

- Además se caracteriza por tener un gasto bajo (20€-50€) y un número de visitas medio con lo que el gasto por visita es reducido

9.5.2.8 SEGMENTO 7 “Sensibilidad muy alta en perfumería”

Tamaño: 8.89%

9.5.2.8. Caracterización segmento 7



	Segmento 7	Global
Nivel redención	4,05%	5,49%
Gasto mensual	38,1 €	93,2 €
Visitas mensual	2,9	4,8
Gasto por visita	17,4 €	22,7 €
Num.grandes categorías	1,5	3,8

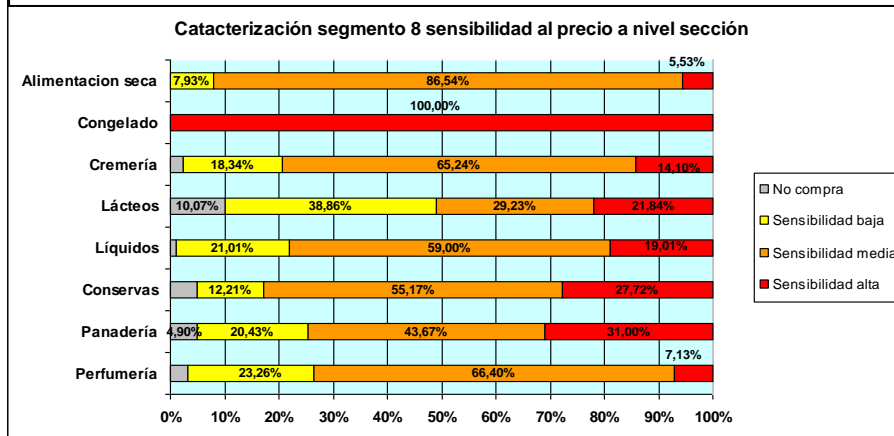
El segmento 7, que representa el 8,89% de todos los clientes, destaca por lo siguiente:

- Todos los clientes de este segmento compran productos baratos en la sección perfumería.
- Además tiene sensibilidad alta en conservas y cremería.
- Para la sección congelado: la mitad no compra y la otra mitad tiene en promedio poca sensibilidad.
- Cuenta con un nivel de sensibilidad media en el resto de secciones.
- Tiene un nivel de sensibilidad a los cupones medio.
- El segmento se caracteriza por tener un gasto bajo (20€-50€) y un número de visitas medio.

9.5.2.9 SEGMENTO 8 “Sensibilidad muy alta en congelado y media en las demás secciones ”

Tamaño: 11.22 %

9.5.2.9. Caracterización segmento 8



	Segmento 8	Global
Nivel redención	5,00%	5,49%
Gasto mensual	92,0 €	93,2 €
Visitas mensual	5,4	4,8
Gasto por visita	22,9 €	22,7 €
Num.grandes categorías	3,7	3,8

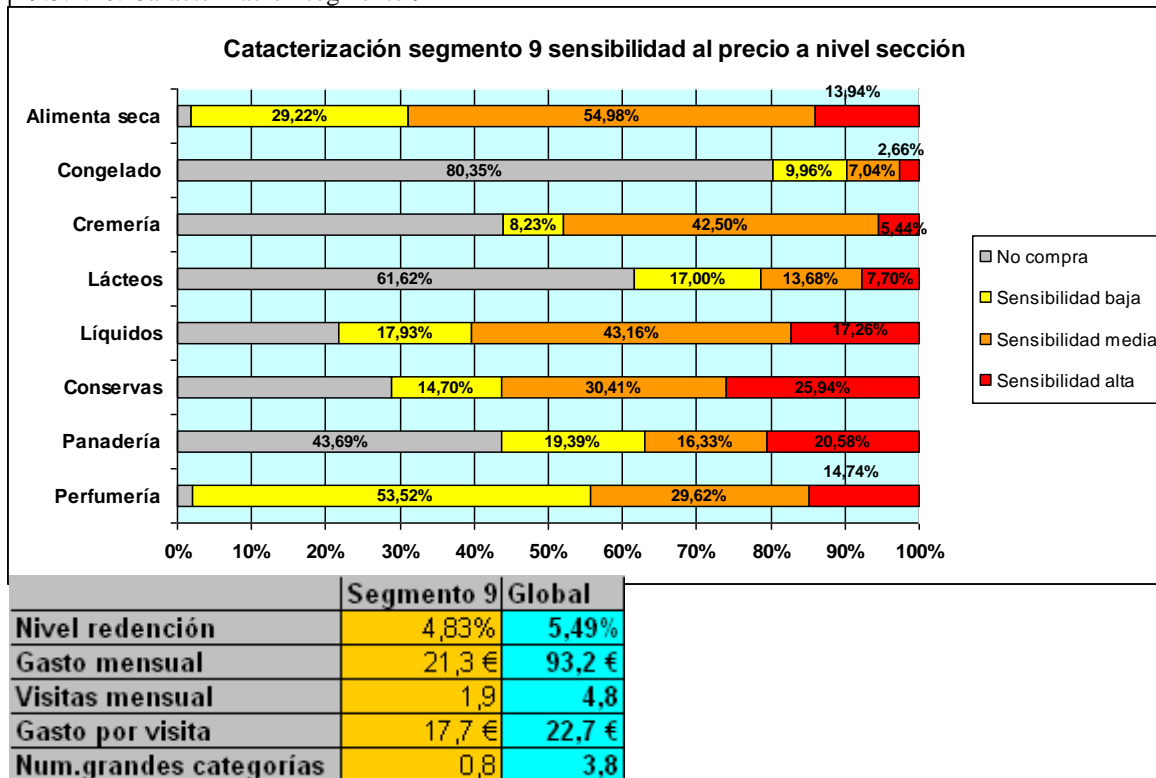
El segmento 8, que cuenta con un tamaño del 11,22% de todos los clientes, se caracteriza por lo siguiente:

- Todos los clientes de este segmento tienden a comprar productos baratos en la sección congelado.
- Para las demás secciones se observa un nivel de sensibilidad medio.
- Es un segmento caracterizado por un gasto mensual y un número de visitas medio. Este segmento tiene un gasto por visita más elevado que los anteriores.
- Tiene un nivel de sensibilidad media a los cupones.

9.5.2.10 SEGMENTO 9 “El segundo grupo que menos gasto realiza”

Tamaño: 7.53 %

9.5.2.10. Caracterización segmento 9



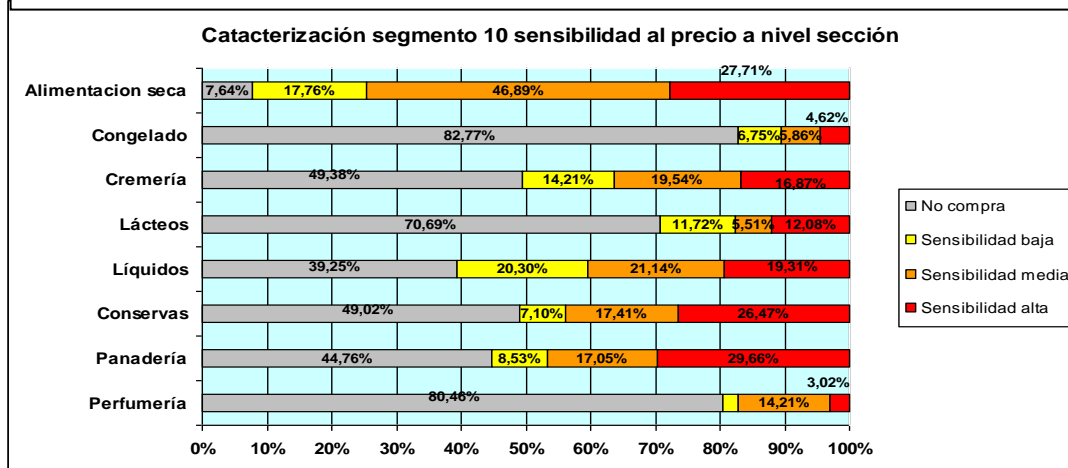
El segmento 9 tiene un tamaño del 7,53% de todos los clientes y se caracteriza fundamentalmente por su escaso nivel de fidelidad:

- Se caracterizan por comprar entre 2 y 10 productos al mes.
- Tienen un gasto y un número de visitas muy reducido, siendo el segundo grupo que menos gasto realiza.
- El 80% de los clientes que componen el segmento no compran en la sección congelado.
- Es un grupo que compra en las secciones alimentación seca y perfumería. Para esta última se observa una cierta tendencia a adquirir productos caros de la sección perfumería.

9.5.2.11 SEGMENTO 10: "El segmento menos fiel"

Tamaño: 5.63 %

9.5.2.11. Caracterización segmento 10



	Segmento 10	Global
Nivel redención	2,98%	5,49%
Gasto mensual	16,3 €	93,2 €
Visitas mensual	1,3	4,8
Gasto por visita	14,1 €	22,7 €
Num.grandes categorías	0,4	3,8

El segmento 10 que cuenta con un tamaño del 5,63% de todos los clientes viene identificado por lo siguiente:

- Es el segmento con menor grado de fidelidad: el que menos gasto mensual y número de visitas realiza.

- El índice de sensibilidad para estos clientes no es muy significativo ya que han comprado pocos productos en secciones: entre 2 y 10 productos al mes.

10. MARKET BASKET ANALYSIS

10.1 ¿QUÉ ES EL MARKET BASKET ANALYSIS?

El Market Basket Analysis, también conocido como análisis de afinidad de productos o Descubrimiento de reglas de asociación (Association rule discovery), es un método de Data Mining que consiste en la identificación de objetos que tienden a ocurrir juntos dado un evento o registro. De una forma más aplicada al negocio de un supermercado el algoritmo del Market Basket Analysis proporcionará un conjunto de reglas de asociación entre productos que tienden a ser comprados juntos. Por lo tanto estas reglas de asociación nos revelan afinidades o correlaciones entre los productos.

El objetivo de este análisis es determinar la fuerza o la potencia de todas las reglas de asociación a lo largo de un conjunto de compras.

El aspecto positivo de esta técnica es la claridad y la facilidad para interpretar y entender las reglas de asociación, por ejemplo, es muy intuitivo y bastante rápido de analizar que si un cliente compra zapatos, entonces el 10% de las veces también compra calcetines. (Linoff&Berry, 2004)

UTILIDAD: APLICACIONES REALES

El objetivo deseado que se pretende con este algoritmo es descubrir asociaciones entre productos que aporten valor añadido a la compañía para la toma de decisiones. La detección de asociaciones entre productos permite apoyar decisiones como:

- Política de descuentos.** Se pueden aplicar políticas agresivas de promociones para aquellos productos que están muy relacionados entre sí. Por ejemplo, si suponemos que existe una fuerte asociación entre productos complementarios (zapatos y calcetines), es decir, se conoce que los zapatos y calcetines tienden a comprarse juntos, los responsables del establecimiento no pondrán ambos artículos en promoción al mismo tiempo. Seguramente la promoción de uno de los artículos propulsará las ventas del otro.

- Ubicación de productos dependiendo de las relaciones entre productos.**

- Creación de packs promocionales.** Una posible estrategia de marketing es crear paquetes o bolsas en los que se ofrezca dos productos con una alta asociación entre ellos más otro producto que no esté tan relacionado y tenga un valor más alto. De esta forma la asociación de dos productos puede incentivar la compra de otro producto más caro.

MEDIDAS DE EVALUACIÓN DE LAS REGLAS DE DECISION E INTERPRETACIÓN

Para realizar un análisis de la cesta de la compra es necesario conocer ciertas medidas que nos permitan evaluar cada regla de asociación. A continuación se exponen las siguientes medidas:

-Support (Soporte): Nos mide el porcentaje de los clientes que compran varios productos juntos respecto a todos los clientes.

Para la regla $A \Rightarrow B$ el Support se calcula de la siguiente forma:

$$\frac{\text{Número de clientes que han comprado } A \text{ y } B}{\text{Número de clientes totales}}$$

El Support es una medida reflexiva, es decir, para la regla $A \Rightarrow B$ el support es el mismo que para la regla $B \Rightarrow A$.

-Confidence: Nos muestra la fuerza de la asociación y mide el porcentaje de que un evento suceda dada la condición de que tenga lugar otro evento.

Según (Linoff&Berry, 2004) es el ratio del número de transacciones que cumplen la regla entre el número de transacciones que cumplen la parte condicional de la regla.

Suponiendo la regla de asociación $A \Rightarrow B$ el nivel de confianza es una probabilidad condicionada:

$$\frac{\text{Número de clientes que han comprado } A \text{ y } B}{\text{Número de clientes que han comprado } A}.$$

-Expected Confidence: Es el porcentaje de clientes que han comprado los productos de la relación en el consecuente entre el número de clientes totales. Más formalmente, es igual al número de transacciones que existen en el consecuente dividido entre el número total de transacciones.

-Lift: Es la confianza dividida entre la confianza esperada, asumiendo que los productos son independientes. Se trata de una medida general de asociación entre conjuntos de objetos o productos.

$$\frac{\text{Confidence}}{\text{Expected Confidence}}$$

- Valores mayores de 1 indican una correlación positiva.
- Valores similares a 1 muestran que no existe correlación.
- Valores menores de 1 señalan una correlación negativa.

El lift es una medida reflexiva, es decir, el lift de la regla $A \Rightarrow B$ es el mismo que el lift de la regla $B \Rightarrow A$.

Si el $\text{lift}=2$ para la regla $A \Rightarrow B$ esto implica que un cliente habiendo adquirido el producto A es dos veces más probable de adquirir el producto B que un cliente elegido al azar. (Bay, 2007).

INTERPRETACIÓN DE LAS REGLAS

En la siguiente figura se muestra un ejemplo de una regla de asociación descubierta en los datos.

Figura 10.1.1. Regla de decisión Alcachofas⇒ Plátanos						
Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
2	33,13	62,45	15,34	1,88	16.917	ALCACHOFAS ==> PLATANO

Según se desprende de la tabla, el 62,5% de los que compran alcachofas han adquirido plátanos también (Confidence). Además existen 16.917 clientes lo que representa el 15,4% de todos los clientes, que han comprado alcachofas y plátanos (Support). Se espera que el 33,13% de los que han comprado alcachofas compren plátanos también (Expected confidence). De hecho el 62,5% de los que compraron alcachofas y plátanos han adquirido plátano. Por lo tanto el Lift será igual a

$$\frac{62,5\%}{33,13\%} = 1,88$$

El Lift representa el 88% de incremento en coocurrencias observadas sobre coocurrencias esperadas.

Figura 10.1.2. Regla de decisión con más de dos productos.						
Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
4	33,13	83,79	1,66	2,53	1.835	TOMATE MADURO & PERAS CONFERENCE & FRESON HUELVA ==> PLATANO

En el caso de reglas de decisión formadas por más de dos productos como en la figura 10.1.2 la interpretación es similar. En este caso el 83,79% de los que compraron tomate, peras y fresas han adquirido plátanos también. Existen 1,835 clientes, lo que representa el 1,66% sobre el total, que han comprado tomates, peras, fresas y plátanos.

10.2 OBJETIVOS

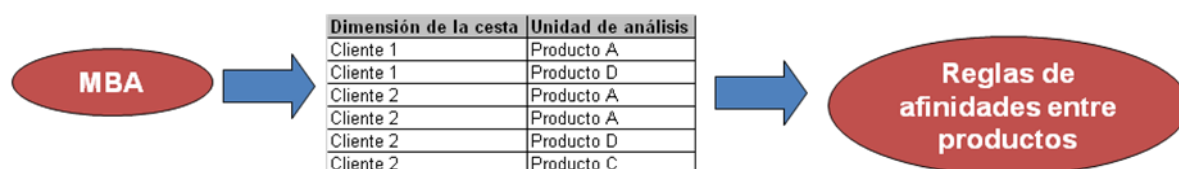
El objetivo propuesto en este proyecto es analizar las diferentes asociaciones y reglas de decisión entre productos aplicando el algoritmo del Market Basket Analysis.

No se trata de un objetivo puntual y concreto sino que se pretende tener una visión general sobre las relaciones de afinidad entre productos. Esta información será útil para un proyecto real de asignación de cupones de descuento a clientes.

10.3 APLICACIÓN EN LOS DATOS

En este apartado se mostrarán los resultados de la aplicación del Market Basket Analysis para las compras efectuadas en el mes de marzo de 2010. En total se trata de más de 5 millones de transacciones correspondientes a las compras de 109,263 clientes.

Figura 10.3.1. Aplicación del algoritmo de Market Basket Analysis.



Para aplicar el algoritmo transformamos los datos tal como se muestra en la figura anterior de tal forma que en la dimensión de la cesta identificamos a los clientes con un identificador único y en la variable unidad de análisis se encuentran los productos adquiridos por cada cliente.

La aplicación del Market Basket Analysis ha sido realizado a través del Software **SAS ENTERPRISE MINER** que tiene implementado el llamado **algoritmo A priori**.

10.3.1 PRIMEROS RESULTADOS

Figura 10.3.1.1. Asociaciones entre productos de la sección frutería.

Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
2	33.13	69.99	7.73	2.11	8.518	PERAS CONFERENCE IMPORTACION ==> PLATANO
2	33.13	69.84	3.97	2.11	4.381	PERA BLANQUILLA ==> PLATANO
2	33.13	69.26	8.03	2.09	8.854	PERAS CONFERENCE GIRONA ==> PLATANO
2	33.13	68.20	3.96	2.06	4.369	ZANAHORIA MANOJO ==> PLATANO
2	33.13	68.01	9.17	2.05	10.108	NARANJA MEJOR MESA G ==> PLATANO
2	33.13	67.94	8.42	2.05	9.284	NARANJA GRANEL MESA ==> PLATANO
2	33.13	66.83	8.29	2.02	9.145	KIWI ==> PLATANO
2	33.13	66.42	5.97	2.00	6.578	MANZANA GOLDEN ESPECIAL ==> PLATANO
2	33.13	65.91	1.99	1.99	2.194	NABOS ==> PLATANO
2	10.22	65.82	1.99	6.44	2.191	NABOS ==> PUERROS
2	33.13	65.28	10.52	1.97	11.604	FRESON HUELVA ==> PLATANO
2	33.13	63.07	9.90	1.90	10.913	PIMIENTO ROJO ==> PLATANO
2	33.13	62.92	6.74	1.90	7.436	TOMATE VERDE ==> PLATANO
2	33.13	62.80	3.95	1.90	4.359	NARANJA BOLSA 2KG ==> PLATANO
2	33.13	62.45	15.34	1.88	16.917	ALCACHOFAS ==> PLATANO
2	33.13	62.15	7.25	1.88	7.990	TOMATE RAMA ==> PLATANO
Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
4	33.13	83.79	1.66	2.529	1.835	TOMATE MADURO & PERAS CONFERENCE IMPORTACION & FRESON HUELVA ==> PLATANO
4	33.13	82.57	1.72	2.492	1.899	TOMATE MADURO & PERAS CONFERENCE GIRONA & NARANJA GRANEL MESA ==> PLATANO
4	33.13	82.50	1.70	2.490	1.872	NARANJA GRANEL MESA & ALCACHOFAS & AJOS SECOS BOLSA ==> PLATANO
4	33.13	82.47	2.07	2.489	2.281	NARANJA MEJOR MESA G & KIWI & ALCACHOFAS ==> PLATANO
4	33.13	81.70	2.03	2.466	2.237	PERAS CONFERENCE GIRONA & FRESON HUELVA & ALCACHOFAS ==> PLATANO
4	33.13	81.70	2.13	2.466	2.344	PERAS CONFERENCE GIRONA & NARANJA MEJOR MESA G & ALCACHOFAS ==> PLATANO

Tras aplicar el algoritmo a todas las compras efectuadas en el mes de marzo de 2010 se obtienen más de 3,000 reglas de asociación. Las reglas de asociación con mayor nivel de asociación (Confidence) tienen en común que están relacionadas con productos que pertenecen a la sección frutería y verduras. Así, analizando la primera regla se deduce que casi el 70% de los que han comprado Peras han adquirido Plátanos.

Como se observa en la tabla 10.3.1.1 ordenada de mayor a menor nivel de confianza (Confidence), el producto plátano aparece en el consecuente de casi todas las reglas con mayor nivel de asociación.

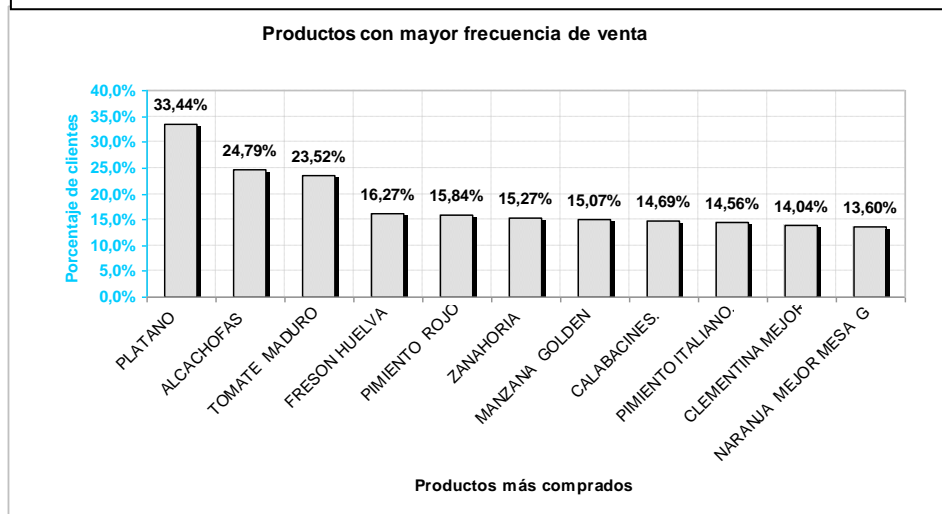
En figura 10.3.1.2 se presentan las asociaciones entre productos de diferentes categorías ordenados de mayor a menor fuerza de asociación (Confidence). Se aprecia lo comentado anteriormente: todas las relaciones tienen en común el consecuente: la fruta plátano.

Esto es debido a que el plátano es el producto del supermercado con mayor éxito y que más veces ha sido adquirido.

Figura 10.3.1.2. Asociaciones entre productos de diferentes categorías.

ASOCIACIONES ENTRE CATEGORÍAS DIFERENTES						
2	26.75	48.99	1.42	1.83	1.580	POLLO SUELTO.TROCEADO. ==> PLATANO
2	26.75	48.89	1.76	1.83	1.969	MUSLO POLLO GRANEL. ==> PLATANO
2	26.75	47.76	2.69	1.79	3.001	HUEVOS MARCA X 6 UNIDADES ==> PLATANO
2	26.75	46.80	1.39	1.75	1.555	PAN RALLADO BOLSA 200 GR. ==> PLATANO
2	26.75	46.79	1.47	1.75	1.640	JAMÓN COCIDO ==> PLATANO
2	26.75	46.61	1.94	1.74	2.159	ARROZ 1K. ==> PLATANO
2	26.75	46.23	2.47	1.73	2.754	JAMON COCIDO EXTRA CAMPOFRIJO ==> PLATANO
2	26.75	45.67	1.39	1.71	1.556	PASTA MARCA X 500GR ==> PLATANO
2	26.75	45.46	1.83	1.70	2.039	YOGURT ==> PLATANO
2	26.75	45.28	1.94	1.69	2.165	ACEITE DE OLIVA CARBONELL 1 L ==> PLATANO
2	26.75	45.11	1.51	1.69	1.683	SERVILLETAS==> PLATANO
2	26.75	45.06	2.15	1.68	2.400	ACEITE DE OLIVA ==> PLATANO
2	26.75	44.58	2.55	1.67	2.840	CERDO LOMO ==> PLATANO

Figura 10.3.1.3. Productos con mayor frecuencia de ventas.



En el gráfico anterior se muestran los productos más comprados por los clientes. El 33,44% de los clientes han adquirido plátanos, el 24,7% alcachofas y el 23,53% tomate maduro.

Los resultados obtenidos hasta el momento no son muy significativos de las relaciones entre productos ya que casi todas las asociaciones encontradas están formadas por aquellos productos con mayor frecuencia de compra (plátanos, alcachofas, tomate,...). Por lo tanto, para detectar más asociaciones entre los productos se procede a eliminar aquellos productos con mayor frecuencia de compra. Más concretamente se procede a eliminar aquellos productos que aparecen en el gráfico anterior para intentar descubrir otras asociaciones que puedan ser interesantes.

10.3.2 RESULTADOS POSTERIORES

Después de eliminar aquellos productos que tenían mayor frecuencia aplicamos el algoritmo a los productos restantes. Aunque se han obtenido más de 1,000 reglas de asociación, en las siguientes tablas se presenta una parte representativa de las reglas obtenidas.

Figura 10.3.2.1. Asociaciones entre productos de la misma categoría.

Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
2	10.23	68.49	1.32	6.70	1.452	CHIRIVIAS ==> PUERROS
2	10.23	65.82	1.99	6.44	2.191	NABOS ==> PUERROS
2	15.15	61.08	1.18	4.03	1.295	CHIRIVIAS ==> ZANAHORIA
2	10.23	60.31	2.33	5.90	2.567	APIO ==> PUERROS
2	15.15	58.52	1.77	3.86	1.948	NABOS ==> ZANAHORIA
2	15.15	58.41	2.26	3.86	2.486	APIO ==> ZANAHORIA
2	13.49	46.98	5.83	3.48	6.420	NARANJA TIPO A ==> NARANJA TIPO B
2	13.48	57.36	0.82	4.25	908	PAN BAGUET INTEGRAL 200G A ==> BAGUETTE DEL HORN A 210 GR.
2	13.48	55.82	1.21	4.14	1.329	PAN BASTON B 252G. ==> BAGUETTE DEL HORN A 210 GR.
2	13.48	46.39	2.11	3.44	2.325	PAN GALLEGO 240GR. A ==> BAGUETTE DEL HORN A 210 GR.
2	5.56	47.24	1.10	8.50	1.216	SALCHICHA DE CERDO FRESCA ==> CERDO LOMO
2	5.56	47.16	0.80	8.49	879	BUTIFARRA DE CERDO FRESCA ==> CERDO LOMO
2	5.56	42.68	1.88	7.68	2.068	PECHUGA POLLO GRANEL ==> CERDO LOMO
2	5.56	42.03	0.82	7.56	907	CAP DE MORT S/PEIXET TERNERA ==> CERDO LOMO
2	4.63	41.98	0.78	9.07	864	CERDO ESPINAZO ==> CERDO COSTILLA
2	5.56	40.66	1.88	7.32	2.073	CERDO COSTILLA ==> CERDO LOMO
2	5.56	40.24	1.71	7.24	1.878	CERDO CUELLOS ==> CERDO LOMO
2	11.43	45.91	2.13	4.02	2.350	ESPARRAGO MARCA BLANCA 8/12 540 G ==> ACEITUNAS MARCA BLANCA 150 G
2	12.76	44.90	1.24	3.52	1.364	LEVADURA 4 SOBRES 64 G ==> AZUCAR BOLSA 1KG
2	11.43	54.68	1.68	4.78	1.851	ALMENDRAS MARCA BLANCA 125G ==> ACEITUNAS MARCA BLANCA 150 G
2	12.76	47.23	0.82	3.70	902	HARINA Y BOLSA 1 KG ==> AZUCAR BOLSA 1KG
2	12.76	46.53	0.78	3.65	857	HARINA REPOSTERIA 1 KG. ==> AZUCAR BOLSA 1KG
2	3.54	39.74	1.00	11.24	1.106	ALUBIAS COCIDAS MARCA BLANCA 540 GR ==> GARBANZOS COCIDOS MARCA BLANCA 540G
2	11.65	20.30	2.64	1.74	2.852	AZUCAR BOLSA 1KG ==> ACEITUNAS MARCA BLANCA 150 G

Analizando la tabla anterior se puede ver que existe una fuerte asociación entre aquellos productos que se encuentran en la misma sección. Los productos pertenecientes a la sección frutería se asocian entre ellos teniendo mayor nivel de Confidence que otros productos de otras secciones.

Así se ve que hay asociaciones importantes entre los productos que pertenecen a la sección carnicería, frutería y panadería.

Además se puede extraer que los clientes tienden a comprar habitualmente productos sustitutos como por ejemplo:

- Pan Gallego y Baguette.
- Carne de cerdo y de ternera.
- Refresco de Naranja y Refresco Coca-cola.

También existe una asociación entre aquellos productos que se consideran complementarios:

- Harina y Azúcar.
- Levadura y Azúcar.

Figura 10.3.2.2. Asociaciones entre productos de diferentes categorías.

Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
2	11,43	54,68	1,68	4,78	1.851	ALMENDRA MARCA BLANCA 125G ==> ACEITUNAS MARCA BLANCA 150 G
2	11,43	45,91	2,13	4,02	2.350	ESPARRAGOS MARCA BLANCA 8/12 540 G ==> OLIVES GEGANTS FARCI SORLI 150 G
2	12,76	37,55	1,23	2,94	1.351	ARROZ 1K. ==> AZUCAR 1KG
2	15,15	35,89	0,93	2,37	1.026	TERNERA HUESOS ==> ZANAHORIA
2	15,15	32,97	1,11	2,18	1.224	MUSLO POLLO GRANEL. ==> ZANAHORIA
2	12,76	35,24	0,77	2,76	846	PASTA MARCA A 250GR. ==> AZUCAR BOLSA 1KG
2	12,76	33,84	0,98	2,65	1.083	PASTA MARCA A500GR ==> AZUCAR BOLSA 1KG
2	12,76	33,05	0,78	2,59	854	SPAGUETTI-3 MARCA A 500 GR ==> AZUCAR BOLSA 1KG
2	12,76	31,91	0,81	2,50	897	PASTA MARCA Z 250 GR. ==> AZUCAR BOLSA 1KG
2	12,76	33,05	1,07	2,59	1.176	GALLETA 800 GR ==> AZUCAR BOLSA 1KG
2	12,76	31,56	1,73	2,47	1.903	ACEITE DE OLIVA 1 L ==> AZUCAR BOLSA 1KG
2	12,76	30,95	1,47	2,43	1.622	ACEITE DE OLIVA MARCA BLANCA 1L ==> AZUCAR BOLSA 1KG
2	12,76	31,49	1,10	2,47	1.207	SERVILLETAS 30X30 1C 200 UTS ==> AZUCAR BOLSA 1KG
2	13,48	31,48	1,24	2,33	1.362	HUEVOS 12U. ==> BAGUETTE DEL HORN A 210 GR.
2	12,76	30,61	1,06	2,40	1.172	LECHE MARCA AENTERA BRIK 1L. ==> AZUCAR BOLSA 1KG
2	15,15	30,60	1,57	2,02	1.731	ESPARRAGO VERDE ==> ZANAHORIA

En la tabla anterior se muestran algunas asociaciones de productos que pertenecen a diferentes categorías.

La de mayor Confidence es la relación de almendras con aceitunas. El 54,68% de los que compran almendras también han comprado aceitunas. Además esta asociación tiene un soporte del 1,68%. Se ha estudiado que esta asociación no se repite en otros meses.

Otras asociaciones descubiertas pero que cuentan con un nivel de confianza inferior son las siguientes:

- Arroz ⇒ Azúcar
- Carne ⇒ Zanahoria
- Pasta ⇒ Azúcar
- Galletas ⇒ Azúcar
- Aceite ⇒ Azúcar
- Servilleta ⇒ Aceitunas
- Leche ⇒ Azúcar.
- Esparrago ⇒ Zanahoria

Estas reglas de asociación aparecen en otros meses.

10.4 CARACTERIZACIÓN DE REGLAS DE DECISIÓN

En este apartado se entrará en detalle en la regla *Plátano* \Rightarrow *Alcachofas* obtenida anteriormente.

Se analizará y se caracterizará a los 16,917 clientes que han comprado Plátanos y alcachofas.

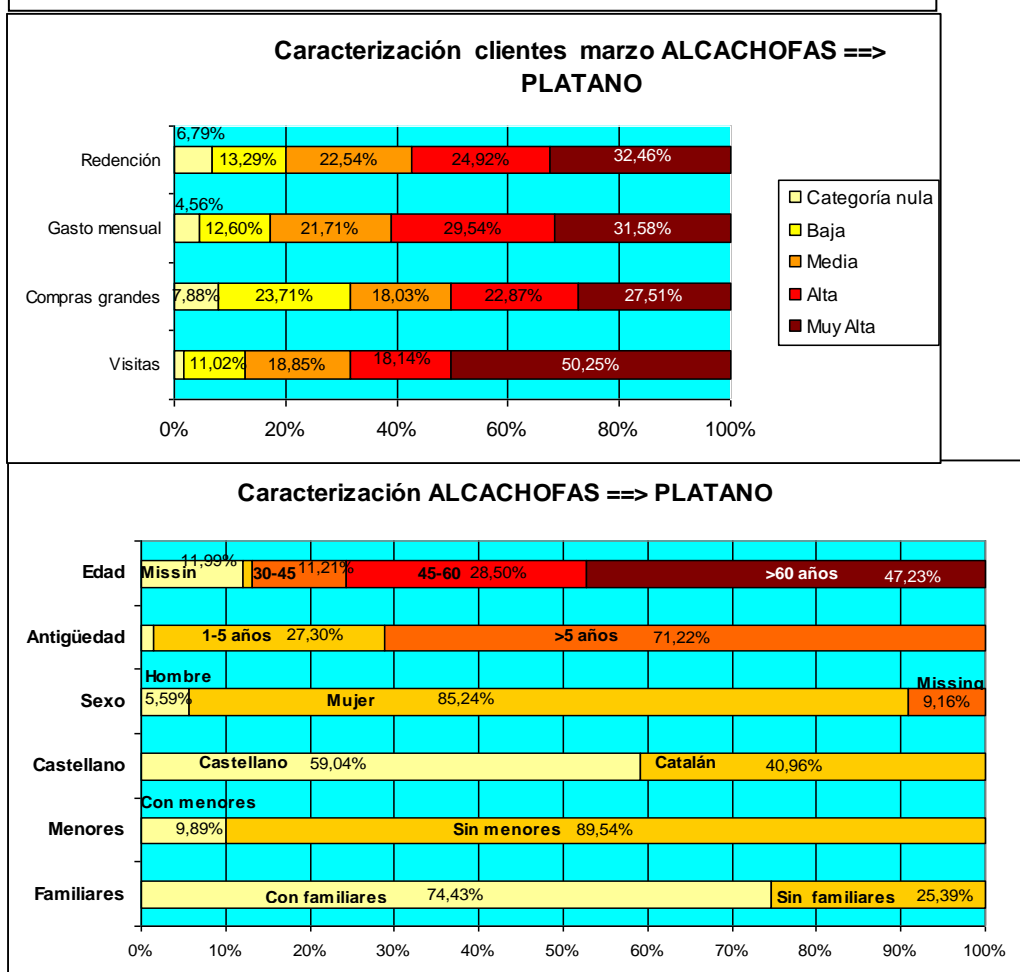


Figura 10.4.1.

Set Size	Expected Confidence	Confidence	Support	Lift	Count	Rule
2	33,13	62,45	15,34	1,88	16.917	ALCACHOFAS ==> PLATANO

A continuación presentamos los gráficos de caracterización de la asociación alcachofas plátanos según las variables de fidelidad y las variables socio-demográficas.

Gráfico 10.4.2. Caracterización de la regla *Plátano* \Rightarrow *Alcachofas*

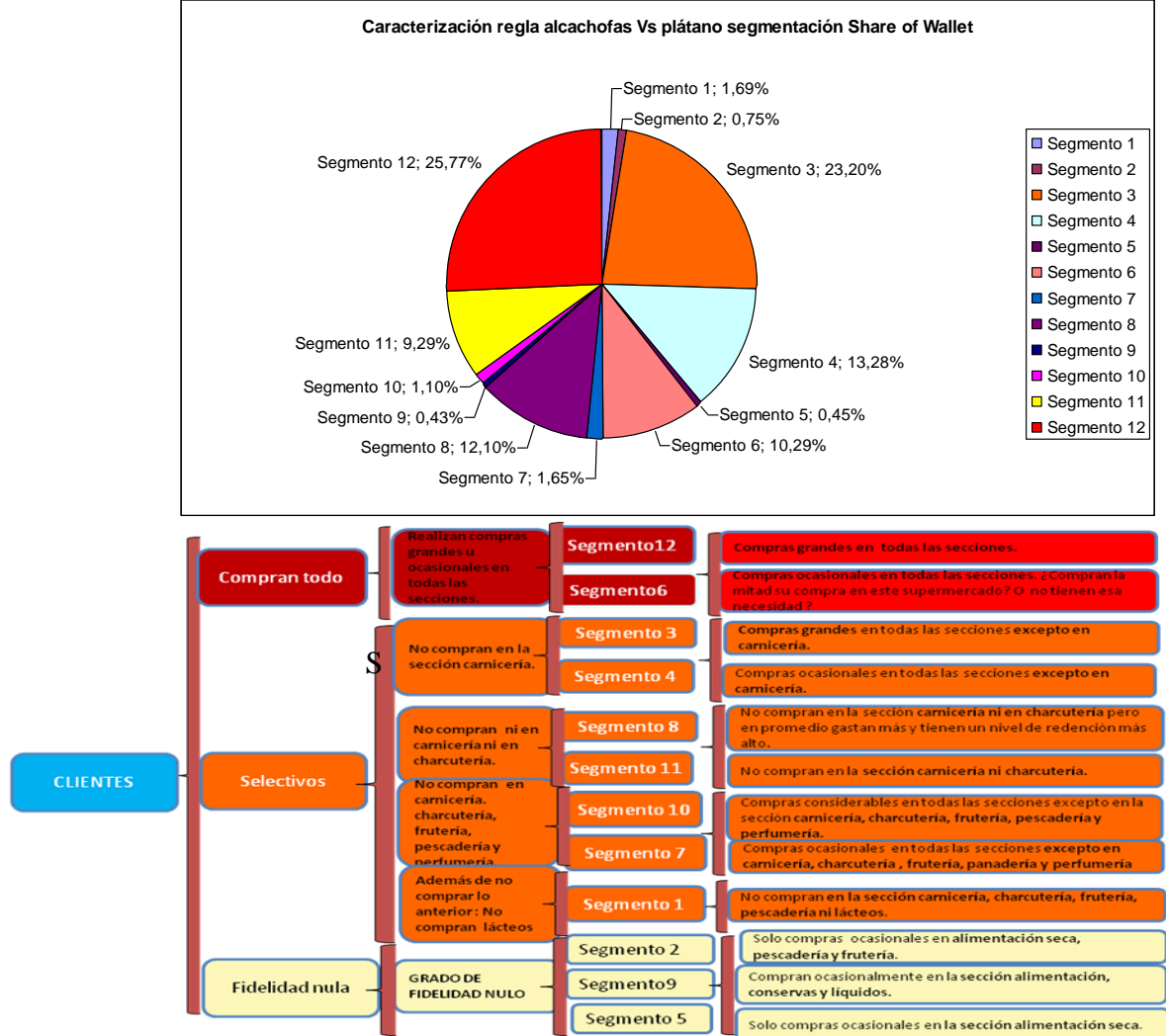


Analizando los gráficos anteriores se extrae el siguiente conocimiento:

- Se trata de un grupo de clientes con una edad elevada (el 47 % son mayores de 60 años).
- Realizan muchas visitas (el 68% realiza más de 5 visitas al mes)
- Son muy sensibles al cupón (el 65% tiene un nivel de redención alta: mayor del 6%).
- Se caracterizan por un gasto elevado: el 60% gasta más de 100€ mensuales en promedio.

A continuación, aprovechando la segmentación del Share Of Wallet realizada en el apartado 8, se observa cómo se reparten los 16.391 clientes entre los 12 segmentos de la segmentación del Share of Wallet.

Gráfico 10.4.3. Caracterización regla alcachofas plátanos según Share of Wallet.



Analizando el gráfico anterior se aprecia que la mayoría de los clientes se encuadran en aquellos segmentos que tienen un mayor gasto y mayor número de visitas y que suelen comprar casi todo en el supermercado (segmento 12 y segmento 3). También se concentran en otros segmentos en los que predomina las compras ocasionales (segmento 6 y segmento 4).

10.5 CONCLUSIONES DEL MBA

Aunque la salida que proporciona el algoritmo es fácil de interpretar y muy intuitiva existen algunos inconvenientes como el excesivo número de reglas de asociación encontradas lo que hace difícil el análisis. De todas las reglas de asociación descubiertas la mayoría se pueden considerar como información trivial, es decir, información que no aporta valor añadido especial a los gerentes del negocio. Se obtienen asociaciones entre productos que son triviales, es decir, son resultados conocidos por cualquier persona familiar con el negocio.

Por ejemplo decir que hay una fuerte asociación entre peras y plátanos o entre levadura y azúcar entra dentro de lo esperado. Además se han detectado algunas reglas que parece que no tienen explicación y no sugieren ningún curso de acción.

No se ha obtenido ninguna regla inesperada que aporte valor añadido y que pueda ser útil para realizar alguna recomendación como la tan famosa asociación entre la cerveza y los pañales que describen Berry y Linof en su libro.

Un pequeño inconveniente de este algoritmo es que es costoso computacionalmente cuando se trata de grandes volúmenes de datos. Aunque SAS es una herramienta muy potente computacionalmente hablando, el tiempo que tarda el algoritmo para mostrar los resultados es de 50 minutos aproximadamente.

En el análisis que se ha presentado se ha intentado detectar qué productos son más probables de aparecer juntos en una cesta de la compra de un cliente pero también sería interesante examinar el orden en el que los productos son comprados. Esto puede ser útil para responder a la siguiente pregunta: si un cliente compró el producto A en la primera visita, cuál es la confianza de adquirir el producto B en la siguiente visita. Esta versión del Market Basket Analysis se conoce con el nombre de Sequence Analysis y requiere la especificación de una variable que marque una secuencia en la compra. El análisis de asociación simple ignora la secuencia o la dimensión temporal de la compra. En este proyecto de fin de máster este análisis no ha sido incluido en la memoria pero ha sido un análisis que ha sido desarrollado en Neometrics.

Bajo mi punto de vista esta técnica de Data Mining es un recurso que puede ser útil para empresas del sector del “retail” pero no hay que esperar de esta técnica el descubrimiento de reglas milagrosas del tipo pañales \Rightarrow cervezas.

11.CONCLUSIONES GENERALES

La identificación de segmentos de clientes sensibles al precio, sensibles a los cupones y con un nivel de fidelidad determinado resulta una información muy útil para el departamento de Marketing de cualquier supermercado. Por ejemplo, la detección de un grupo de clientes que son sensibles al precio en unas secciones determinadas e insensibles en otras puede ser una herramienta muy potente para desarrollar un plan de fidelización y personalizar acciones de Marketing.

Además una segmentación de los clientes en función del Share of Wallet puede resultar clave para la aplicación de políticas comerciales personalizadas de tal forma que fomenten el gasto en aquellas secciones en las que un segmento de clientes no compra.

También hay que decir que un análisis de la cesta de la compra puede permitir un conocimiento que ayude a tomar decisiones referentes a la ubicación de productos y a la mejora de una política promocional.

VALORACIÓN PERSONAL DEL PROYECTO

Este proyecto ha supuesto una colaboración con la empresa Neometrics desde octubre del 2010 hasta julio de 2011. Se trata de un proyecto complementario que sirve de apoyo a un proyecto real que está realizando Neometrics en la actualidad. Desde el punto de vista profesional estoy satisfecho de haberme enfrentado a un proyecto analítico real de Minería de datos y de trabajar con datos reales. Especialmente, he tenido que manejar un gran volumen de datos. Se ha trabajado con 180 millones de registros de compras pertenecientes a 200,000 clientes por lo que ha sido necesario el aprendizaje de lenguaje de bases de datos relacionales (SQL). Además he aprendido a manejar la herramienta estadística SAS (ENTERPRISE MINER y ENTERPRISE GUIDE) realizando macros para automatizar ciertas operaciones repetitivas.

Además es importante afirmar que este proyecto ha supuesto mi primera experiencia profesional en el mundo del análisis de datos o Data Mining y también he tenido la oportunidad de colaborar en un ambiente internacional y multidisciplinar. Yo soy licenciado en Administración y Dirección de Empresas y en Investigación y Técnicas de Mercado, y el hecho de trabajar con personas de diferentes titulaciones (Matemáticos, Informáticos, Telecomunicaciones, Ingenieros de Caminos...) supone que el aprendizaje sea continuo y enriquece el ambiente de trabajo.

Finalmente he tenido la ocasión de aplicar ciertos conocimientos adquiridos en las asignaturas del MEIO, especialmente me han sido de gran utilidad asignaturas como Métodos Multivariantes, Técnicas Cuantitativas de Marketing y Minería de Datos ya que me han proporcionado un conocimiento general e indispensable para el desarrollo de este proyecto.

CONCLUSIONES DEL ESTUDIO

Se ha visto que las variables de fidelidad gasto mensual, número de visitas y número de compras grandes están asociadas entre sí de una forma lineal pero la variable redención se escapa de esa tendencia. Además aquellos segmentos identificados por una alta sensibilidad a los cupones se caracterizan por tener un gasto por visita muy reducido y un gasto mensual más bajo comparado con aquellos segmentos que tienen un nivel de redención medio. El segmento de los “amantes de los cupones” se caracteriza además por ser el segmento de mayor edad y por no realizar todas sus compras en el mismo establecimiento.

Además se ha descubierto que aquellos segmentos con un mayor nivel de fidelidad se caracterizan por un nivel de redención medio, por tener una antigüedad superior a 5 años, una edad comprendida entre 45-60 años, por vivir en el hogar con familiares y menores de edad.

Respecto al análisis del “Share of Wallet” se aprecia una tendencia por parte de los clientes de no comprar productos perecederos en el supermercado. Da la sensación de que las familias prefieren comprar la carne y el pescado en otros establecimientos.

Después de realizar las segmentaciones en función de la sensibilidad al precio se ha descubierto que niveles reducidos de redención se asocian con niveles bajos de sensibilidad al precio mientras que niveles altos de redención están más asociados con niveles altos de sensibilidad al precio. Lo más destacado es que clientes con un nivel medio de sensibilidad al precio tienden a tener un nivel de fidelidad alto mientras que aquellos clientes que tienen cierta preferencia por adquirir productos baratos tienen un nivel de fidelidad más reducido.

Finalmente, analizando las asociaciones que nos proporciona el Market Basket Analysis la mayor parte de las asociaciones encontradas se consideran como triviales, es decir, son resultados conocidos por cualquier persona familiar con el negocio. A pesar de esto se ha detectado que hay asociaciones importantes entre productos que pertenecen a la misma categoría, destacando la sección frutería. Además se ha detectado que los clientes tienden a comprar productos sustitutos (carne de cerdo y de ternera) y hay asociación entre ciertos productos complementarios (levadura y azúcar).

Bibliografia

- [1] **Escofier B & Pagès J. (1992). Análisis factoriales simples y múltiples: objetivos, métodos e interpretación.** Edición Española. Servicio editorial de la Universidad del País Vasco.
- [2] **Tomá Aluja Banet (1996). Análisis Factoriales Descriptivos con SPAD-N.** Centre publicacions d'Abast
- [3] **Aluja T. & Morineau A (1999). Aprender de los datos: El análisis en componentes principales. Una aproximación desde el data mining.** Ediciones Universidad de Barcelona.
- [4] **Tomás Aluja Banet (2001). La minería de datos entre la estadística y la inteligencia artificial.** Quaderns d'Estadística, Sistemes, Informatica i Investigació Operativa, ISSN 0210-8054, Vol. 25, Nº. 3, 2001 , págs. 479-498.
- [5] **Michael J. Berry & Gordon Linoff . Data Mining Techniques: For Marketing, Sales, and Customer Support.** John Wiley & Sons, Inc. New York, NY, USA.
- [6] **Xinli Bai (2007). Mining Transactions/Order Data using SAS Enterprise Miner: Association Node.** SAS GLOBAL FORUM 2007.
- [7] **Linoff & Berry, (2004). Gordon S. Linoff, and Michael J. Berry. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Support.** John Wiley & Sons, Inc. New York, NY, USA.
- [8] **Matthew Redlon.(2003) “A SAS Market Basket Analysis Macro: The 'Poor Man's Recommendation Engine'.** Proceedings of the twenty-eighth Annual SAS Users Group International Conference. April 2003.
- [9] **SAS Enterprise Miner Help Menu.**
- [10] **SAS Enterprise Guide Help Menu.**
- [11] **Luis Aburto Lafourcade. “Beneficios del Market Basket Analysis: Experiencias en un Supermercado de Santiago”.** Ingeniería Industrial. Universidad de Chile. Enero 2011.
- [12] **L. M. Gherghi & C. Lauro. “Appunti di analisi dei dati multidimensionali: Metodologie ed esempi”.** Università di Napoli Federico II. Dipartimento di Matematica e Statistica, Naples, 1999.

